

The hand that feeds

The efficiency of research agencies and their responsiveness to grant applicants vary widely around the world. It is time for the laggards to reform.

The integrity of the grant-selection process should be the central objective of every research agency — and scientists who apply for grants fully understand that the process must be thorough and exhaustive.

Nonetheless, as we report in this issue (see page 308), grant applicants confront an array of service quality ranging from first-rate to abysmal. Like other consumers of government services, they are entitled to expect a satisfactory level of service from the agencies that they deal with — including politeness from staff, timely responses to enquiries, and ready access to all non-confidential information pertaining to the grant review process.

Instead, some researchers report interminable paperwork, unexplained delays, arbitrary rules and regulations, and ultimate bafflement regarding the actual nature of the review process. Some of this is probably sour grapes from failed grant applicants. And grumpy researchers should bear in mind that the agency's overriding obligation is to the taxpayer whose money it dispenses. Nonetheless, many researchers' complaints are justified, and reflect a level of public-sector incompetence that must be urgently addressed by the agencies in question.

Most of these agencies argue, with varying degrees of persuasiveness, that they are working to improve things. But pinning down their progress can be difficult. Some agencies are tracking their performances against indicators such as the amount of time taken to acknowledge correspondence or to reach decisions. But these criteria — and agencies' performances against them — are often hard for grantees or other taxpayers to track down. And if no one knows the benchmarks, it is hard for anyone to be held accountable.

Some of funders' most useful innovations have been borrowed from the private sector, such as the use of surveys and focus groups

to get feedback from grantees. But the analogy with private industry cannot be pushed too far. Research agencies are public bodies with public responsibilities. On the other hand, they face no real competition: researchers can't choose between funders in the same way that consumers can between suppliers.

Pressure on the public sector to improve its performance can move research agencies in the right direction. In Japan, for example, the government of Prime Minister Junichiro Koizumi has pushed for greater transparency and efficiency from its research agencies. In the United States, the Bush administration has asked federal agencies to streamline their submission systems (www.grants.gov), a move that could help the best practices of the strongest research agencies, the National Science Foundation and the National Institutes of Health, to rub off on the rest of government.

But other political pressures can cause bureaucratic demands to stack up. Concerns over the treatment of minorities and women, for example, or over financial conflicts of interest, often lead to additional requirements on grant application forms — which are always easier to add than to remove. Agencies need to look at their forms over time, to see if they still make sense. Researchers, for their part, should bear in mind that the forms may reflect legal requirements over which the agency has no control, as well as priorities, such as fair treatment of applicants, that the researchers probably share.

It is particularly important that agency staff have time to get out and meet the communities they serve: direct contact with these staff is particularly valuable for young scientists who are new to the application process. And discussions with researchers suggest that their level of satisfaction with agencies can be improved through relatively simple steps, such as responding quickly to complaints and maintaining a telephone helpdesk that is manned by real people. ■

Global reach

Nations need a more effective way to coordinate their responses to environmental challenges.

The United Nations Environmental Programme (UNEP) will be looking for a new director-general at the end of this year, and the UN World Summit this week may consider steps to widen the body's influence.

UNEP's departing director, former German environment minister Klaus Töpfer, has worked hard during his eight years in office to strengthen the body's ties with industry, and to get its work taken seriously beyond the narrow circle of environmental groups that have supported its efforts in the past.

But a great deal more needs to be done if this relatively obscure office, with about 450 professional staff and an annual budget of US\$60 million, is to make any real impact on epic global environmental problems such as global warming, clean air, clean water and biodiversity conservation.

UNEP was set up in 1972 with a mission “to provide leadership and encourage partnership in caring for the environment by inspiring, informing, and enabling nations and peoples to improve their quality of life without compromising that of future generations”.

The programme has scored some notable successes on Töpfer's watch, helping to coordinate international discussions on water-supply issues, for example, and helping poor countries develop laws and regulations on complex issues such as the transportation of biological specimens and transgenic plants.

But UNEP is not currently constituted to provide genuine leader-



ship on critical environmental issues, even when these cry out for an international response. Unlike fully fledged UN agencies such as the Food and Agriculture Organization or the World Health Organization, it is merely a 'programme', funded on a voluntary basis by the United Nations' member states.

Perhaps the outfit whose clout contrasts most vividly with UNEP's is the World Trade Organization, which is independent of the United Nations and arbitrates forcefully and effectively in disputes between its member states. UNEP, on the other hand, has neither ways to settle disputes nor mechanisms to enforce compliance with international environmental agreements such as the Convention on Biological Diversity.

UNEP's remit is more modest than that. Its job is to set and monitor standards for environmental protection and sustainable development around the world, in collaboration with local governments, scientists, non-governmental organizations and other interested parties. This responsibility stretches from biodiversity to climate change, from managing clean water to desertification, and from bio-safety to problems posed by invasive species.

Under Töpfer, UNEP has improved its use of scientific information, and gained better access to the corridors of corporate power. On the downside, Töpfer has been drawn into distracting public disputes with other international bodies, over issues such as the administration of the Global Environment Facility, which finances environmental projects (see *Nature* 394, 4; 1998).

But UNEP's real problem is that it lacks the power to enforce the growing number of binding environmental agreements between nations. Beefing up the programme would probably involve a mandatory funding scheme based on the size of members'

economies. Plans for this have been around for years, but they face significant obstacles, starting with the opposition of the United States, which currently contributes less than either the United Kingdom or Germany to UNEP's budget.

Such funding concerns hide a broader fear, by no means confined to the United States, that a more powerful UNEP would constrain the freedom that national governments currently enjoy to pollute pretty much as they please. This may be short-sighted, however. In the long run, national governments — and global capitalism, for that matter — might benefit from a strong international environmental body, a World Environmental Organization, if you will, with a remit to safeguard the future of the planet.

Töpfer has helped UNEP to build bridges with the worlds of business and finance, and

sought to convince business leaders that sustainable development and a healthy environment are in their interest, too.

The UN Summit could strengthen UNEP in the short term by merging it with the smaller, separate UN Division for Sustainable Development, currently a branch of the UN Department of Economic and Social Affairs. Secretary-General Kofi Annan then needs to appoint a heavyweight successor to Töpfer who can provide UNEP with energetic and determined leadership. That person should continue the policy of partnership with industry, while carefully guarding the organization's independence and further nurturing its credibility, in preparation for the day when national governments are ready to upgrade its status. ■

"UNEP's problem is that it lacks the power to enforce the growing number of binding environmental deals between nations."

All things equal

Lack of affordable child care is a major impediment to women's careers, in science as elsewhere.

The problem of under-representation of women in science, particularly at the most senior levels, is not going to go away. Public discussion of the issue often focuses on the extent to which girls are encouraged to pursue scientific interests at school, or to which they are discriminated against at work. But a more readily addressable impediment obstructs the career paths of many female researchers in early- and mid-career: the absence of suitable child-care arrangements.

The issue of child care tends to arise at a crucial juncture in women scientists' careers, and there is a growing consensus that it can play a significant role in thwarting scientific ambitions. All over the world, even as the number of women who pursue graduate education continues to grow, women remain under-represented in senior scientific positions. Among the major scientific nations, the situation is perhaps most acute in Japan and in Germany, where women make up 30% of those starting graduate school, but only 6% of full professors.

Advocacy groups such as the Association for Women in Science frequently emphasize the need to encourage girls to do science at school, and to mentor women scientists early in their careers. They should place equal emphasis on the need for affordable day care. Without it, women scientists can be forced to choose between putting off having children, or having their careers derailed by motherhood.

Most major universities and laboratories offer some child-care options, but uneven access and affordability prevent anything resembling a working, national system in all but a few countries. These programmes vary widely in quality and even the best ones have significant shortcomings, as we report in this issue (see page 446). Scandinavian nations and France offer some of the most comprehensive arrangements, but these are stretched thinly, with long waiting lists the norm. Most US universities offer child care, but the cost can be prohibitive to junior researchers.

Some employers have sought to address the affordability issue: the European Molecular Biology Laboratory in Heidelberg, Germany, for example, charges a fee of 10% of parents' combined income. Individual senior researchers have also chipped in to tackle the problem: Nobel laureate Christiane Nusslein-Volhard, head of the Max Planck Institute for Developmental Biology in Tübingen, set up a foundation last year to give five women scientists at the laboratory €400 (US\$490) a month for babysitting and domestic help, when she noticed that talented female research assistants were dropping out of science once they had children.

The only systems that really work, though, involve government subsidies or tax incentives that enable the considerable cost of child care to be spread between the state, the employer and the employed. If women's representation in science is truly their objective, governments and research institutions must find a way to share the child-care load. ■

"Uneven access and affordability prevent a working, national child-care system in all but a few countries."

RESEARCH HIGHLIGHTS

Skeleton clocks

Cell **122**, 803–815 (2005)

The genes that make the body's circadian clock tick are also involved in controlling bone formation, according to a team led by Gerard Karsenty at Baylor College of Medicine in Houston, Texas. The findings offer avenues for tackling bone diseases such as osteoporosis in which bone density (pictured) decreases.

Bone formation has a daily rhythm, so Karsenty's team studied mice that had been genetically engineered to lack key clock genes. They found that such mice had heavier bones because they had more bone-building cells. The clock-gene proteins respond to signals from the hormone leptin and control the proliferation of these cells. Intriguingly, bone seems to be the only tissue where leptin acts through this pathway, suggesting this is an evolutionarily ancient and important aspect of leptin biology.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

C. MUNOZ-YAGUE/EURELIOS/SPL

CHEMICAL BIOLOGY

Genomic miners

Nature Chem. Biol. doi:10.1038/nchembio731 (2005)

Finding natural compounds that have important medicinal properties has always been a laborious process of searching and purification.

Now, chemists have a powerful tool to speed up their hunts. A team at the University of Warwick, UK, has used the genome sequence of a bacterium to predict that it makes a particular molecule.

The team, led by Gregory Challis, studied DNA sequences from *Streptomyces coelicolor* (pictured below) and identified clusters of genes involved in biosynthesis. Using these, the group predicted the existence of an unknown molecule, which it then isolated from the bacteria and named coelichelin. This technique, which has been dubbed

'genome mining', could be used to discover other natural products.

GENETICS

BRCA2 risk for men

J. Med. Gen. **42**, 711–719 (2005)

The *BRCA2* gene, which has been linked to hereditary breast cancer in women, is associated with an increased risk of prostate and pancreatic cancers in men, says a Dutch study.

Flora van Leeuwen of the Netherlands Cancer Institute in Amsterdam and her colleagues investigated 139 families with 66 different mutations of the *BRCA2* gene. They studied the incidence of cancers among family members who had a 50% chance of carrying the mutant gene. The team compared these incidences with those expected in the general population and found increased risks of pancreatic and prostate cancers in men under the age of 65. Male *BRCA2* carriers may also have a greater risk of bone and throat cancers.

The enzymes studied are normally mobile and control the addition of ubiquitin to HIF (hypoxia-inducible factor) and p53, a tumour-suppressor protein. But when temporarily captured in the nucleolus, the enzymes are denied access to these molecules. The researchers propose that cells have evolved a mechanism to regulate the function of proteins by switching the enzymes between mobile and static states.

NEUROBIOLOGY

Wrap artist

Neuron **47**, 681–694 (2005)

A gene known as *neuregulin* helps to provide the insulating coat that certain nerve cells require. James Salzer of the New York University School of Medicine and his co-workers show that switching on this gene triggers the production of myelin. This fatty material envelopes some neurons and allows them to send signals up to 100 times faster than their bare counterparts. The team studied mouse neurons that are normally myelinated but had been engineered to lack both copies of the *neuregulin-1 type III* gene; these cells remained bare. The researchers also discovered that neurons that are usually unmyelinated acquire the protective layer when altered to express the gene.

CELL BIOLOGY

Bound and gagged

J. Cell Biol. **170**, 733–744 (2005)

Cells can control the activity of enzymes by locking them up in a nuclear structure called the nucleolus, say researchers in Canada.

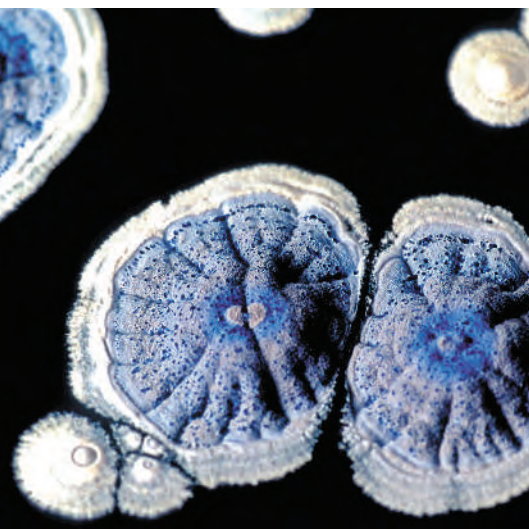
Stephen Lee and colleagues from the University of Ottawa, Ontario, studied two enzymes called ubiquitin ligases. When active, these attach a molecule called ubiquitin to proteins. This alters a protein's destiny, often signalling its destruction.

STEM CELLS

Keep your options open

Cell **122**, 1–10 (2005)

One big question about embryonic stem cells is how they maintain their potential to form



any cell type in the body — an ability known as pluripotency. Now, a team headed by Richard Young at the Massachusetts Institute of Technology sheds light on the cellular control circuits involved.

The researchers studied three proteins called OCT4, SOX2 and NANOG, which control gene activity and are known to be important for pluripotency. Working with human stem cells, the team scanned the whole genome and identified the genes controlled by these proteins.

In many cases, the three proteins acted together on the same genes, activating those involved in pluripotency and division, and repressing those that direct embryonic development.

IMMUNOLOGY

Barrier method

Nature Immunol. **6**, 995–1001 (2005)

A newly discovered 'firewall' that protects our cells against viral invasion could be exploited to develop antiviral drugs, say US researchers.

Roughly half of the viruses that cause human disease are surrounded by a membrane. These invaders, which include flu viruses and HIV, need to fuse their membranes with those of their host cells to enter them.

A team led by Leonid Chernomordik at the National Institutes of Health in Bethesda, Maryland, has found that a set of peptides called defensins, which form part of our innate immune system, can block this fusion. In tests on cultured human cells, they showed that defensins reversibly link cell surface proteins together, forming a barrier that stops the membranes getting close enough to fuse.

CANCER

A sticky situation

Nature Genet. doi:10.1038/ng1635 (2005)

A variant of a gene called *Sipa1* seems to play a key role in helping tumours to spread. Kent Hunter of the National Cancer Institute in Bethesda, Maryland, and his colleagues have found that damping the expression of *Sipa1* significantly slows tumour spread in mice. And engineering tumour cells to make extra *Sipa1* protein doubles the chance that they grow and disperse. The team shows this extra protein can cause a cell to lose its adhesive properties, perhaps freeing it to colonize other organs. Tumours that have spread from human prostate cancers also express abnormally high levels of *Sipa1*.

PLANETARY SCIENCE

Dust buster

Science doi:10.1126/science.1118923 (2005)

On 4 July 2005, NASA's Deep Impact mission slammed a probe into Tempel 1 to find out what lies in a comet's interior. The resulting series of papers reveal detailed quantitative data on the comet's composition and structure. In an overview paper, Michael A'Hearn of the University of Maryland, College Park, and his colleagues describe how Tempel 1 is covered in impact craters, which have never been seen before on comets. The surface of Tempel 1 has surprising features, with both old- and young-looking terrains, and signs of past geological processes. The comet consists of very fine, loose particles, and its interior contains organic compounds.



GEOMORPHOLOGY

Chillout dunes

Geology **33**, 753–756 (2005)

The last place most people would look for an ice sheet, such as the one pictured above, is the Sahara Desert. But that is where Julien Moreau of the School and Observatory of Earth Sciences in Strasbourg and his colleagues have found one. It isn't there any more, naturally; it existed about 440 million years ago, when Africa was part of the supercontinent of Gondwana situated over the South Pole.

Moreau and his team have found signatures of glaciation — scratches and ridges in rocks and elongated deposits called drumlins — near the town of Ghat on the border of Algeria and Libya.

These features were made by an ice stream, a fast-flowing section of an ice sheet, and they are the oldest clear evidence of such streams yet found.

M. VAN WOERT/NOAA NESDIS/ORA

JOURNAL CLUB

Frank Wilczek

Massachusetts Institute of Technology, Cambridge, USA

The promise that anyon particles hold for quantum computing excites the physicist who named them.

When I first worked on — and named — anyons in the early 1980s, I thought these particles, whose quantum behaviour goes beyond the familiar categories of bosons and fermions, were an interesting

theoretical curiosity.

To my amazement, anyons soon appeared in the theory of the quantum Hall effect, which describes the odd modifications of electronics that occur at low temperatures in strong magnetic fields. But relevant experiments are difficult, and have been a long time coming.

Now Fernando Camino of Stony Brook University in New York and colleagues report observations of peculiarities in the quantum Hall effect that they interpret as the first experimental indication of anyon

behaviour (<http://www.arxiv.org/cond-mat/0504341>; 2005). In a separate paper (<http://www.arxiv.org/cond-mat/0412343>; 2005), Sankar Das Sarma from the University of Maryland and his colleagues propose a refined experimental arrangement, which could be used both to clarify this behaviour and to make it useful for quantum computing.

Anyons become useful for quantum computing when they wind around each other, so that their world lines form tangles in space-time. The different tangles represent

states in Hilbert space, the vast arena of quantum mechanics, and can be used to store information. Hilbert space is big enough to accommodate very complicated tangles, so huge storage and bandwidth could be in the offing.

The theory of anyons has spawned some beautiful but specialized and difficult mathematics, hitherto pursued only by a small cult. I suspect that many of my colleagues have, like me, been waiting for a pretext to dive back in. With these two papers, we have it.

NEWS

New Orleans researchers fight to salvage work from submerged labs

Hurricane Katrina has apparently devastated research laboratories in New Orleans. Rescue teams last week discovered that many frozen specimens and cell cultures had thawed, making them useless, and laboratory animals had drowned.

"This was a real disaster for research," says pharmacologist Joseph Moerschbaeher, vice chancellor for academic affairs at the Health Science Center of Louisiana State University (LSU) in New Orleans.

Academic rescue teams may have saved some projects, entering flooded buildings and topping up dewars of liquid nitrogen to keep specimens cold. But the full scientific loss may not be known for weeks or months.

The catastrophe may trigger debate over how prepared academic institutions ought to be for natural disasters. Four years ago, a storm destroyed similar facilities at a medical center in Houston (see *Nature* **411**, 874; 2001).

With more than half a dozen colleges and universities effectively out of commission along the US Gulf Coast, scientists and students have begun scattering across the country to try to salvage their careers and studies. Dozens of institutions opened their doors last week to displaced researchers, who had often lost their homes as well as their bench space. Scientists based in Louisiana or Mississippi have moved as far as California to maintain their projects.

The National Science Foundation, the National Institutes of Health (NIH), and the Department of Energy are extending grant deadlines, offering to match affected researchers with academic hosts, and generally doing what they can to help along recovery from Katrina.

Late last week, Katrina's flood waters began to recede in New Orleans as the US Army Corps of Engineers patched breaks in the city-protecting levees and started pumping the water back into the neighbouring Lake Pontchartrain. On 6 and 7 September, more than a week after Katrina

hit the region, researchers made their first trips into the medical and research complexes of LSU and Tulane University in downtown New Orleans. Accompanied by rescue workers and armed military personnel, small teams assessed the damage to labs that had been underwater or without power for many days.

Trying to keep cool

At the LSU centre, the flood water was too high for a Humvee to navigate, so the team climbed aboard trucks, says pharmacologist Wayne Backes, LSU's assistant dean for research and lead academic on the survey team. Once inside, the researchers found that a dozen buildings were flooded with two metres of water.

The team walked the dark hallways with flashlights, searching for research material that could be saved before they had to evacuate by the nightly curfew. "It was hot, humid and smelled like mouldy refrigerators," Backes says. "All the -70°C freezers were at room temperature. A lot of samples were destroyed."

The team searched for liquid-nitrogen dewars to replenish, particularly those containing transgenic cell lines. "It was very difficult to find things in labs we were not familiar with," Backes says. "We pulled out stuff as best we could."

LSU's four animal research facilities, containing rodents, dogs and primates among others, were on the flooded ground floors. "100% of those lab animals were lost," says Moerschbaeher.

At the Tulane centre, the animals were located on the eighth floor, well beyond the reach of the flood waters. A survey team there also discovered freezers without electricity, but didn't have the chance to fully explore because of the high waters.

John Clements, head of microbiology and immunology at Tulane, said the 6 September expedition was so harried that he could not check on his own home, just three blocks off the truck caravan's path. Tulane's main campus, located on higher ground uptown, survived Katrina in relatively good shape.

The Katrina disaster will raise questions about how scientific institutions should prepare for storms and who should take responsibility for emergency measures, especially given past experiences in the region. In 2001, the tropical storm Allison swamped the Texas Medical Center in Houston, drowning research animals and destroying years of cell lines.

Together LSU and Tulane have about 300 researchers who are annually funded by \$130 million of NIH money. Shortly after Katrina hit, NIH director Elias Zerhouni said it was not his agency's responsibility to help universities weather catastrophic storms. That task, he argued, belonged to the Federal Emergency Management Agency. But as the agency struggles to feed and house tens of thousands of people caught in Katrina's path, research is a low priority. ■

Rex Dalton



Residents were still struggling to leave as scientists headed for their labs.

W. BACKES

**HURRICANES IN FOCUS**

Access an archive of all our stories on extreme storms.

www.nature.com/news/infocus/hurricanes.html

B. AHO-NAYI/UP/NEWS.COM

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Dirty job: the water is being pumped out of New Orleans, but residues of petroleum, pesticides, garbage and sewage are expected to remain.

First tests show flood waters high in bacteria and lead

As the clean-up operation along the US Gulf Coast gets into full swing, the environmental effects of Hurricane Katrina are slowly coming to light.

Samples of the flood waters swamping New Orleans taken from residential areas between 3 September and 5 September have revealed dangerous levels of lead and coliform bacteria. The US Environmental Protection Agency (EPA) has promised a thorough analysis, and has told people to avoid contact with the waters if possible.

Along with the expected sewage contamination, the waters in New Orleans contain a wide range of chemicals. Some 150,000 submerged cars may have leaked petrol, oil and antifreeze, and an equal number of houses may have contributed asbestos from insulation and lead from paint. Unknown quantities of household pesticides, cleansers and solvents are also likely to be in the mix.

The flooding submerged the Agriculture Street Landfill, a former rubbish dump that,

once covered by a protective liner, has been developed into a residential area. No one knows yet whether any of the toxic waste in the landfill has begun leaching through the liner.

"We know the water is bad," EPA administrator Stephen Johnson said last week. The agency has tested for more than 100 chemicals in the flood waters, including volatile organic compounds, pesticides, herbicides and polychlorinated biphenyls.

As of early this week, the only chemical on that list reported to be at dangerously high levels was lead in one residential area. And in a dozen samples taken over three days, all but two swamped the maximum detection level of 2,400 colonies of *Escherichia coli* per sample. The recommended limit is 126 colonies.

The Gulf Coast is home to much of the US petrochemical industry; reports of damage to chemical and oil facilities are still emerging. Of more immediate concern are the nearly 1,000 drinking-water systems and more than 100 wastewater systems that have been battered by

the storm, leaving many in the region without safe water. "Nobody has ever dealt with anything on this scale before," says Jean Kelly, a spokeswoman for the Louisiana Department of Environmental Quality, which is working with the EPA.

Agency officials say that the EPA will release a sampling plan for chemical and biological contaminants within days to weeks, using experts from its science board. Until results come through, it is left to other experts to speculate and worry — not only about the city of New Orleans, but about other shattered Gulf Coast communities and the oil refineries and chemical plants along the Mississippi River.

Many of the old wooden homes in New Orleans may have soaked up toxic compounds from contaminated mud, says Neal Langerman of the chemical consulting firm Advanced Chemical Safety in San Diego. "I expect an awful lot of structures will have to be razed," he says.

Emma Marris

US tests satellite tool for hurricane monitoring

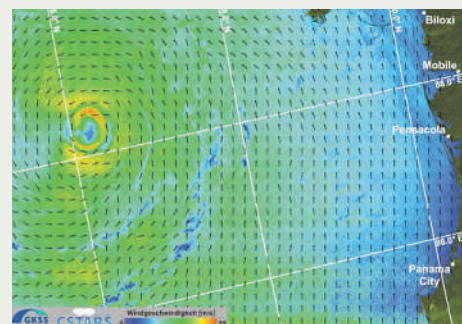
The daredevil aircraft that fly into the hearts of hurricanes, including Katrina, could soon be joined by a satellite tool that can make storm planning more effective.

Synthetic aperture radar (SAR) can monitor wind speeds and directions inside storms at high spatial resolutions, German and US researchers say. The technique, is being tested at the US

National Hurricane Center in Miami.

The new SAR images (see picture) slightly underestimated the storm's maximum wind speed over the ocean, says Jochen Horstmann, an oceanographer at the GKSS research centre in Geesthacht, Germany. The data are from the European ENVISAT and the Canadian RADARSAT-1 satellites, which monitored ocean surface roughness.

Quirin Schiermeier



J. HORSTMANN

Ornithologists stunned by bird collector's deceit

SANTA BARBARA

The British Colonel Richard Meinertzhagen earned many illustrious titles during his life of globe-trotting, including soldier, spy and even scientist. Now an extensive analysis of bird specimens he collected across continents adds another label: fraud artist.

Meinertzhagen, who died in 1967, donated some 20,000 bird specimens to the British Natural History Museum in Tring in 1954. The museum has one of the world's largest ornithological collections, with around two million items, and is an internationally important archive for biology and ecology studies.

But a large proportion of the specimens Meinertzhagen donated were fraudulently labelled and many were stolen, according to a report presented late last month at the annual meeting of the American Ornithologists' Union in Santa Barbara, California.

Pamela Rasmussen, an ornithologist at Michigan State University in Lansing, and Robert Prys-Jones, head of the collection at Tring, began investigating Meinertzhagen's specimens from Asia after fraudulently labelled finches were reported in 1993. It has taken sophisticated forensic techniques and well over a decade of work to uncover the extent of the deceit. "There are hundreds and probably thousands of fraudulently catalogued

"There are hundreds and probably thousands of fraudulently catalogued specimens."

specimens," says Rasmussen. "This was going on for the better part of his life."

A member of the Royal Fusiliers, Meinertzhagen served in the Middle East, Africa and across the Asian subcontinent, campaigning with the likes of Lawrence of Arabia. Substantial portions of his collection are thought to be valid and important, including many specimens from Afghanistan. But Rasmussen and Prys-Jones found that, as early as 1914, Meinertzhagen was stealing specimens from the British Museum and other institutions, then retagging them with catalogue details of his own making. "It's the most egregious example of fraud that I am aware of," says Prys-Jones.

And it may have played havoc with the conclusions of biologists who subsequently studied the specimens. Museum collections of bird skins, skeletons, nests and eggs are the primary record by which biologists chart the life of species. Stored items are marked with small tags that contain crucial details of a specimen's life history, such as favoured climate and geographic location.

Rasmussen and Prys-Jones's studies included painstaking analysis of the way specimens were prepared and the handwriting on tags, together with X-rays of skeletons, to determine which had faked data, and to match specimens with those stolen from other institutions. They

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Detailed, accurate tagging of museum specimens is crucial for biologists researching bird species.

found that Meinertzhagen frequently reported species at locations where they had never been seen before, and documented recent sightings of rare or near-extinct birds.

In one case, he took a specimen of the kingfisher *Alcedo hercules* that had been found on the island of Hainan off China, and listed it as being from Myanmar. In another, he swiped a forest owl (*Heteroglaux blewitti*) skin from the British Museum and retagged it as his own

J. BLAIR/CORBIS

Free mice herald launch of Asia-Pacific network

SYDNEY

Developmental biologists in Asia-Pacific countries have come up with an intriguing incentive to boost research in the region — giving away transgenic mice for free.

The service is part of a research network launched in Sydney last week at a meeting of the International Society of Developmental Biologists (ISDB), which aims to encourage collaboration between labs in the region. To kick things off, the RIKEN Center for Developmental Biology (CDB) in Kobe, Japan, will make transgenic mice to order for scientists in the region, free of charge.

Studies using transgenic mice can reveal the functions of genes, particularly during embryonic development. But the mice are expensive to create, so providing them for free should help researchers in developing countries in the region, says Shinichi Aizawa of the CDB, who heads the service.

"The service will be very useful, especially for those scientists who don't have their own facilities to generate knockout mice," agrees Kathy Cheah, a developmental biologist at the University of Hong Kong.

But there are strings attached. The CDB will own the mice, and requires co-

authorship on the first publication generated from the use of each mouse. Ownership will be retained to ensure that the mice produced are available to all scientists, says Aizawa. He adds that if the mice are to be used commercially, an agreement will need to be negotiated between the various parties involved.

The mice will be the first step in bolstering regional ties, says Masatoshi Takeichi, director of the CDB and newly elected president of the ISDB. "Interactions between countries in the region are very weak, and yet geographically we are so close," he says.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

discovery. That gave him credit for having found the last of the species in 1914, when it had actually been discovered in India in the 1880s by another collector. Rasmussen herself has since rediscovered the forest owl alive in India, while she was investigating Meinertzhagen's role in the fraud.

Rasmussen and Prys-Jones have now corrected as many of the tags as possible, and are soon to publish a long article on the malfea-

sance, to serve as a guide for scientists studying the Tring collection. It may also inform the use of other Meinertzhagen specimens, which are housed at the American Museum of Natural History in New York and the World Museum Liverpool, UK. Those collections aren't believed to have such problems, say officials at the museums, but they haven't been thoroughly checked. ■

Rex Dalton

The research network, called the Asia-Pacific Developmental Biology Network, includes Japan, China, Taiwan, Singapore, South Korea, India, Australia and New Zealand, but in the future could extend as far as Iran and Hawaii. It is a grass-roots initiative supported by, but independent of, the ISDB. "The aim is for members to share each other's laboratories with no barriers," says K. VijayRaghavan, director of the National Centre for Biological Sciences in Bangalore, India.

Participants will share ideas and information through the Internet, and host institutes will help cover travel and research costs for students and researchers. Members will also take advantage of each other's

specialist facilities, including the CDB's transgenic mouse production scheme, India's liver-cell imaging technology and Singapore's zebrafish expertise. "The goal is to create a vibrant intellectual environment in the region," says VijayRaghavan.

The research network was lauded by delegates in Sydney. "Researchers in Asia have lots of links with labs in the West, but not so many within the region. We hope the network will change that," says Cheah.

The CDB will start its mouse service this month with collaborators in India. Researchers should receive the mice within a year of sending the DNA constructs or sequence data needed to make the animals. ■
Carina Dennis

Can we score the millennium goals?

NEW YORK

As world leaders gather in New York this week to discuss the future of the United Nations' Millennium Development Goals, critics are warning that many of the targets cannot be evaluated scientifically. Unless the goals are peer reviewed and amended to take account of what can actually be measured, they say, the project will fail.

Nearly 150 heads of state met in 2000 and pledged to fight problems ranging from poverty to infectious disease in a historic document called the Millennium Declaration. This was turned into a series of numerical targets with a final deadline of 2015 that are known as the Millennium Development Goals.

The ambitious and concrete-sounding aims captured imaginations and headlines worldwide, and have become symbolic of the global development effort. They include halving the number of people that live on less than a dollar a day and cutting deaths in children under five by two-thirds.

But analysts such as Amir Attaran of the University of Ottawa, Canada, say that without drastic revision, the targets risk being meaningless. "We haven't got the tools in place to measure them," he says.

The UN Statistics Division currently keeps tabs on 48 'indicators', which range from school enrolment numbers to malaria death rates. These are provided by agencies such as the World Health Organization (WHO) and the World Bank, but Attaran argues that the methods used to gather data are scientifically flawed. In the absence of birth and death registries in most African countries, for example, many figures on death and illness are gathered from household surveys. These often use conflicting methods, and do not always acknowledge sources of error or apply statistical tests of significance.

And although surveys may measure some indicators accurately — such as childhood deaths, which parents will remember — others are harder to estimate, such as cases of malaria or deaths in childbirth. For example, if a woman succumbs to fatal bleeding in hospital after giving birth at home, the link may not be recorded.

Attaran warns, in *PLoS Medicine*, that without accurate baseline figures and ongoing measurements it will be impossible to tell in 2015 whether the millennium goals have achieved anything. He wants independent peer reviewers to assess individual targets and indicators to determine whether they can be measured

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The United Nations has set targets for the relief of human suffering, but progress is hard to measure.

scientifically. If they cannot, he says, the goals should be amended or even abandoned.

Epidemiologists agree that the various indicators could be made more specific, and say leaders should take the opportunity this week to clarify them. "It needs to be talked about now," says Robert Kim-Farley of the University of California, Los Angeles, who used to gather public-health data for the WHO.

"It is relevant to discuss whether we are investing enough in data-collection systems," adds Stefano Bertozzi, director of health economics for the National Institute of Public Health in Cuernavaca, Mexico.

Some UN scientists have expressed concerns like Attaran's. They tried to start a debate last year about how to measure the goals more accurately. But the deputy secretary-general,

Louise Fréchette, deflected them in a message sent to the expert committee in charge of goal statistics. The New York summit "should not be distracted by arguments over the measurement of the millennium development goals — or worse, over the different numbers being used by different agencies for the same indicators", she said.

John McArthur is deputy director of the UN Millennium Project — an independent body based in New York that advises the United Nations on the goals. He agrees that the priority for this meeting is simply for nations to affirm their support. He acknowledges that some of the targets are tough to track, but adds, "I don't think that's a reason not to make a concerted effort."

Roxanne Khamsi

D. TELEMANS/PANOS

Europe backs trials on drugs for kids

LONDON

The European Parliament decided last week that drugs authorized for use in children should first be subjected to clinical trials in appropriate age groups.

Although researchers hope extra trials will lead to a better understanding of how medicines work in children, critics are concerned that patent extensions intended to sweeten the decision could delay the availability of cheap drugs.

Many medications given to children in Europe are not licensed for the situations in which they are used. Doctors often take a trial-and-error approach, simply halving the adult dose, for example. But in some cases such drug use has caused side-effects and even fatalities.

There are several reasons why drug companies are reluctant to test their products in children. It is often necessary to develop new equipment and techniques for



FIT BODY FOR FIT MIND
Aerobic exercise is the best way to keep your memory healthy, says a review of ten years' research.
www.nature.com/news

UK embryo licence draws global attention

Scientists in Britain have been granted permission to perform controversial experiments that will create human embryos using genetic material from three people. Teams carrying out similar research in the United States and China have been forced to shut down their experiments, and say that they hope the UK decision will help push the work forward around the world.

The UK Human Fertilisation and Embryology Authority (HFEA) said on 8 September that it will allow scientists at the University of Newcastle upon Tyne to transfer the nucleus of a fertilized egg into an egg donated by a second woman. The second egg's own nucleus will be removed before the transfer occurs, but it will still contain genetic material in certain structures outside the nucleus. These structures, called mitochondria, generate energy and carry their own genes. So the Newcastle scientists will be creating embryos that contain genes from three people: the biological mother, the biological father and an unrelated egg donor. That makes the experiments highly controversial.

The embryos that will be created in Newcastle will not be transferred into women. But advocates of the technique say that if the experiments work, they could eventually prevent mothers from passing on diseases caused by mitochondrial defects to their children.

Similar experiments were being done in the United States by James Grifo, a reproductive endocrinologist at New York University Medical Center. But in the late 1990s the US Food and Drug Administration (FDA) halted Grifo's experiments after he had created only a few embryos, because it was worried that the health of the fetuses could not be guaranteed. Grifo's postdoctoral fellow, John Zhang, con-

tinued the experiments in China, where altered embryos were transferred into two patients before an international outcry shut those experiments down too. Grifo, who has since been pleading with the FDA to allow him to restart the experiments, told *Nature* that he is encouraged by the HFEA's decision.

"I'm glad that at least one country in the world is pioneering and smart enough to do this, and hopefully it will help patients with mitochondrial diseases," Grifo says.

He predicts that the British move will eventually spur action elsewhere, including in the United States. "Once there's some

semblance of success over there, they'll say 'OK, now we can do it.' Traditionally, that's how things go."

More than 30 children have already been born thanks to a related technique that is used to boost fertility. The method does not involve nuclear transfer — instead, scientists transfer some of the cytoplasm that surrounds a cell's nucleus from a donor egg to the fertilized egg. But that method has also been halted in the United States, leaving the Newcastle researchers at the head of the field, for the moment at least. ■

Erika Check

EDELMANN/SPL

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Egged on:
researchers in Newcastle will be allowed to create an embryo that contains genes from three people.

use on children. And there are sticky ethical issues, such as deciding who gives informed consent when parents are divorced, or what happens when the parents say one thing and the child another.

But the main obstacle, says Sharon Conroy, a paediatric pharmacologist at the University of Nottingham, UK, is the time and money it takes to do the extra tests. Especially as children are relatively healthy, so pharmaceutical companies do not stand to gain much from investing in research on drugs for them.

As an incentive to companies, the

European Parliament and the Council of Ministers are supporting a six-month extension on existing patents, similar to measures offered by the US Food and Drug Administration since 1998.

The European Generic Medicines Association supports the directive in principle, but says the extension may delay the provision of affordable medications for those in financial need.

Mark Schreiner, chairman of the Committees for the Protection of Human Subjects at the Children's Hospital of Philadelphia, points out that the equivalent

regulation in the United States has been of tremendous benefit to research. "We've learned how different children are from adults," he says. For example, Viagra is good for treating lung conditions in children. "We never would have discovered this if drug companies had not been required to carry out rigorous tests with kids," he says.

The parliament's vote is a major step towards enforcing the regulation legally. The proposal was first put forward last September and is expected to become law when the parliament next meets, in 2006. ■
Jennifer Wild

ON THE RECORD

“A bunch of people with strong egos and God complexes. That sounds like rock ‘n’ roll to me.”

Publisher Bob Guccione explains why scientists, the focus of his newly purchased magazine *Discover*, aren't so different from the musicians featured in his former magazine *Spin*.

“We were almost done with our ark and were training the animals to march in two-by-two, but we just didn't make it.”

Dan Maloney of New Orleans' Audubon Zoo explains why staff stayed behind to care for the zoo's animals when Hurricane Katrina hit.

Source: *New York Times*, Reuters

SCORECARD

**Spotted eggs**

It's not decoration, nor is it camouflage. The reason some eggs are speckled comes down to strength. The red spots on eggs laid by great tits occur where the shell is at its thinnest.

**Energy at sea**

Russia unveils plans to build the world's first floating nuclear power plant. The project is already throwing up some unusual safety questions, including how to protect it from attack by divers.

**Bats**

Some 40% of China's horseshoe bats are carrying a SARS-like virus. Bat meat and faeces, used in traditional food and medicine respectively, may now see a drop in demand.

NUMBER CRUNCH

100 billion bases of sequenced genetic code have been deposited in three major public databases since 1982.

165,000 is the number of organisms whose complete or partial genome is coded for by these bases.

90% of this information was deposited during the past five years.

Spacecraft on course to score a first with asteroid samples

Japan's most ambitious space mission entered a final, critical phase this week as the Hayabusa spacecraft parked itself 20 kilometres from the asteroid Itokawa. If all goes to plan, the mission will be the first to return samples of an asteroid to Earth.

The US\$100-million Hayabusa mission is primarily a test of technology, and will carry out “a lot of firsts for the Japanese”, says Andrew Cheng of the Johns Hopkins University Applied Physics Laboratory in Laurel, Maryland. Cheng is one of several Americans on the project's science team.

A Japanese-built ion propulsion system delivered the craft to the 600-metre-long asteroid, and Cheng considers this a success in itself. A strong solar flare in 2003 degraded the spacecraft's solar panels, reducing the power output of the ion drive engines and causing a three-month delay in reaching the asteroid. One of three reaction wheels that help Hayabusa to hold a steady position has also failed, although project officials say the effects on the observing programme will be relatively minor.

More hurdles lie ahead. After mapping the asteroid's surface with cameras, and X-ray and infrared spectrometers, Hayabusa will move in for a closer survey. In November, it will make brief contact with Itokawa, firing tantalum ‘bullets’ into the asteroid, then capturing the material that flies up in a horn-like collecting tube. Scientists hope to gather a gram or more of material from two or three locations.

Just before firing the first bullet, the craft

will drop a tiny ‘hopper’ called Minerva on to the asteroid. This cube, equipped with cameras and thermometers, will flip around the surface like a tiddlywink in the low gravity. Finally, in December, Hayabusa will head home, its samples tucked inside a hermetically sealed capsule that will land near Woomera,

Australia, in 2007.

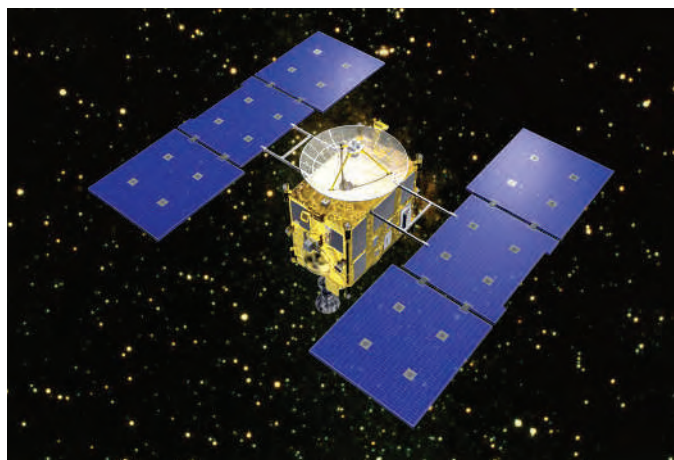
Cheng praises the mission as “very innovative”, and says the scientific pay-off should be high if all the technology works as hoped. No spacecraft has ever visited such a small asteroid. Ground observations suggest that Itokawa, which was only discovered in 1998, is an S-class asteroid similar in make-up to a class of stony meteorites known as chondrites.

The samples from Hayabusa will be compared with telescope observations and meteorites on Earth. So far, scientists have matched only one meteorite to an asteroid source — Vesta — and even that link is debatable, says Michael Zolensky of NASA's Johnson Space Center in Houston, Texas, who will help to analyse the samples that come back. Having a piece of a well-studied object such as Itokawa will give researchers greater confidence in making assumptions about asteroids based solely on spectral analysis from Earth.

Once the samples arrive they will stay in Japan for a year, where a team of Japanese, US and Australian researchers will have first crack at them. The samples will then be shared with other international researchers, and some 10% will be turned over to NASA permanently. ■

Tony Reichardt

“The scientific pay-off should be high if all the technology works as hoped.”



On target: the Japanese craft Hayabusa (left) will collect material from the asteroid Itokawa (above) and return it to Earth.

Incoming ethics chief set to move focus off stem cells

The US President's Council on Bioethics has a new leader — Edmund Pellegrino, an 85-year-old ethicist who is expected to target conflicts of interest in science and medical research.

Observers predict that Pellegrino will take the council in a very different direction from Leon Kass, its previous leader. Kass triggered controversy last year by dismissing a renowned biologist who did not share his views on stem-cell research (see *Nature* 428, 4; 2004). The council has so far played a visible and influential role in US public debate over such research.

Pellegrino, who takes over on 1 October, is an emeritus professor at Georgetown University Medical Center in Washington DC. He is known for his work on whether and why scientists or physicians do the right thing.

"Ed is as good as it gets in bioethics," says Glenn McGee, a bioethicist at Albany Medical College in New York. "He will invite the strongest and best thinkers to debate."

Enthusiast uses Google to reveal Roman ruins

Using satellite images from Google Maps and Google Earth, an Italian computer programmer has stumbled upon the remains of an ancient villa.

Luca Mori was studying maps of the region around his town of Sorbolo, near Parma, when he noticed a prominent, oval, shaded form more than 500 metres long. It was the meander of an ancient river, visible because former watercourses absorb different amounts of moisture from the air than their surroundings do.

His eye was caught by unusual 'rectangular shadows' nearby. Curious, he analysed the image further, and concluded that the lines must represent a buried structure of human origin. Eventually, he



Analysis of a Google map led to the discovery of a Roman villa like this one in Parma, Italy.

CULTURAL MINISTRY/EMILIA ROMAGNA

Basking robot takes the plunge

Solar-powered vehicles that work underwater may sound like a far-out idea, but US robotics experts have made them a reality. This 170-kilogram solar autonomous underwater vehicle, or SAUV, can dive down to 500 metres, says robotics expert Arthur Sanderson of Rensselaer Polytechnic Institute in Troy, New York.

The vehicle surfaces to recharge its batteries using solar panels. These top-ups mean it runs for longer than other subs and can make more measurements of aquatic properties such as levels of dissolved oxygen. Sanderson and his



A. C. SANDERSON/RPI

colleagues envisage several of the vehicles diving in concert, giving three-dimensional glimpses of the underwater world.

traced out what looked like the inner courtyards of a villa.

Mori, who describes the finding on his blog, *Quelli Della Bassa*, contacted archaeologists, including experts at the National Archaeological Museum of Parma. They confirmed the find. At first it was thought to be a Bronze Age village, but an inspection of the site turned up ceramic pieces that indicated it was a Roman villa. The local authorities will have to approve any archaeological digs before they can take place.

► www.quellidellabassa.org

Bird lovers file suit to protect woodpecker

The recent rediscovery of the ivory-billed woodpecker (*Campephilus principalis*) in Arkansas has spawned a lawsuit.

The National Wildlife Federation and its Arkansas chapter want to halt a US\$320-million irrigation project, which they say may harm the bird's habitat. On 8 September, they filed a lawsuit against the US Army Corps of Engineers and the Department of the Interior to stop construction on the Grand Prairie Area Demonstration Project.

The environmentalists allege that the long-running project to divert water for rice farming will damage the woodpecker's habitat along the White and Cache rivers. Federal officials say the project would not harm the bird, reportedly found last year after a half-century hiatus (see *Nature* 437, 188–190; 2005).

University of California toasts first valley campus

A farming city in the sunny San Joaquin Valley is host to the University of California's first new campus in 40 years. The University of California, Merced, which

opened last week, has no departments and puts its graduate students in multi-disciplinary groups covering topics such as environmental systems and quantitative systems biology.

Merced is pitching itself as both interdisciplinary and diverse. A quarter of its first undergraduate class is Hispanic, and almost half will be the first in their families to go to college.

The small founding faculty includes Roland Winston, the solar-power pioneer from the University of Chicago; materials scientist Christopher Viney; and physical chemist Anne Myers Kelley, formerly of Kansas State University. Keith Alley, the vice-chancellor for research, estimates that the campus will draw about US\$8.5 million in external grants next year.

US agency airs policy for pesticide tests on humans

The US Environmental Protection Agency (EPA) last week issued its policy governing studies that intentionally dose people with pesticides (see *Nature* 437, 24–25; 2005).

The proposed rules set scientific and ethical standards that studies must meet if the EPA is to consider them when it is determining safe exposure levels. The policy allows pesticide companies to use data from earlier studies — some dating from the 1960s — provided they met the ethical standards of their time. Past studies of pregnant women and children who happened to be exposed to pesticides will be allowed, but studies that intentionally dosed these groups will not.

In 1998, the EPA effectively banned pesticide manufacturers from using human toxicity studies to get their products to market. But it began reconsidering this on a case-by-case basis in 2003, after manufacturers successfully sued the agency.

The rules are open for public comment for 90 days, and must be finalized by February.



discovered just how petty his country's bureaucracy could be. Last year, he received ¥50 million (US\$450,000) in government funds. Under plans to make Japan's research institutes more independent, the university was entrusted to manage the grant, with some bizarre consequences. When Taira's secretaries asked university officials if they could purchase pens and erasers for the office, they were instructed that such supplies were to be used only in connection with the grant in question. The university denies this, saying it merely assumes that researchers already have pens and would "like them not to buy them again". Either way, absurd accounting wins over common sense.

Arbitrary rules such as these may raise a sigh — but flawed funding decisions can have much more serious consequences. Most scientists need to win competitive grants to keep their careers going. If the application process is mishandled, they can fall behind in the race to publish or, in extreme cases, be forced out of a field altogether.

445

The number of pages in the annual document that describes the budgetary rules associated with grants from Japan's health ministry.

The latter almost happened to Hugh Roberts. Last year, the UK-based immunologist, who asked *Nature* not to use his real name, was facing a tough career decision. During eight years of postdoc work, Roberts had authored well-received papers in top journals. "My CV was as good as anyone's," he says. But he was struggling to make the transition to group leader. Some UK agencies offered career-development grants for this purpose, but the schemes were restricted to applicants with no more than six years' postdoctoral experience. At the age of 34, Roberts felt he was still a young scientist, "but I was being told I was over the hill".

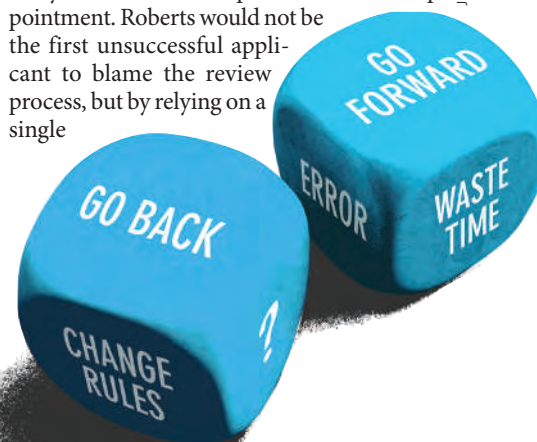
Then a potential solution appeared. The Medical Research Council (MRC), the body that distributes most UK government money for biomedical projects, removed the time limit for one of its grants. Roberts was able to apply for a three-year award of £250,000 (US\$460,000), which he planned to use to set up a lab. The university where he worked was keen to keep him and agreed to pay his salary.

To reduce applicant numbers, the MRC decreed that any university could put forward no more than three people. So Robert's institute set up an internal competition, requiring researchers to submit an eight-page proposal and give a ten-minute talk. He won and submitted an application to the MRC in May. "On and off it was the main thing I was doing for a month," he says.

Roberts was philosophical when told in October that he had not been successful, as he knew there had been more than 120 applications for just 20 grants. "It was tough, but I didn't mind losing in those circumstances," he says. Until, that is, he received his rejection letter from the MRC, and saw the basis on which it had made its decision: a single review from a scientist who, judging by a copy of the review seen by *Nature*, held strong views that were not representative of the field. With no recourse to appeal the decision, Roberts felt let down by the MRC. "They never really gave the grant a chance," says Roberts.

Feeling rejected

Tales of accounting rules and bureaucratic delays seem trivial compared with this disappointment. Roberts would not be the first unsuccessful applicant to blame the review process, but by relying on a single



J. MAGEE

Paper, paper, everywhere...

When *Nature* asked scientists to name the most frustrating grant-application process, one phrase got mentioned again and again: the European Commission's €17.5-billion (US\$21.8-billion) Sixth Framework Programme.

It is, say researchers who have applied, the most confusingly bureaucratic procedure imaginable. A nightmare of mile-high paperwork that requires mental acrobatics to penetrate the Byzantine language and complex eligibility criteria. It may also be beyond major reform, structured as it is around giant collaborations that satisfy the needs of politicians more than bench scientists.

Just working out which grant to apply for is hard enough. A multimillion-euro Integrated Project perhaps? There are five types of integration to choose from, none of them straightforward to understand. 'Vertical integration', for example, is described as involving "the full 'value-chain' of stakeholders from those involved in knowledge production through to technology development and transfer".

Then come reams of forms demanding information that can seem distant from the research project in question. A life-sciences grant, for example, requires an explanation of the gender relevance of the work. "It's hard to project a gender relevance onto research on the basic molecular biology of an asexual microorganism," points out one Dutch researcher.

Equally tortuous is identifying and recruiting the multiple partners needed to make up an international and interdisciplinary collaboration that mixes academics, industrialists and sociologists. All this, and still a coherent scientific research plan to construct.

Why was such a system ever designed? The answer lies with the political reasoning behind the Framework programme. Rather than existing to fund science, it was created to support the policy aims of the European Union (EU). And although all research agencies are linked to political paymasters, the connections between European politicians and the Framework programme are particularly tight.

So the programme is pulled in many different directions. Ministers from EU states, as well as the directly elected European Parliament, have to approve the Framework structure and research themes, for example. Research priorities are also shaped by states pushing their national interests. And the politicians' appetite for breaking down all conceivable barriers requires that the programme plays its part in solving wider social problems — such as gender discrimination.

Following past financial scandals, the commission added further bureaucracy to protect itself from accusations of financial mismanagement. Researchers have to submit detailed finances for each project in advance, and the actual expenditure afterwards. National research grants are usually paid out on



trust and detailed accounting occurs only after the project ends.

Can things improve? Commission bureaucrats recognize the problems but few are optimistic about major changes. The best hope, say scientists, is the planned European Research Council for basic research, a separate funding agency that will be part of the next Framework programme, which starts in 2007. The council is expected to be accompanied by a much lighter bureaucracy and may distribute as much as €1.5 billion annually to individual investigators. But large, unwieldy collaborations will continue to dominate European funding — and to infuriate its researchers.

Alison Abbott

reviewer, instead of the usual three, the MRC seems to have erred. Researchers who sit on grant panels confirm that it is very rare for applications to be dismissed on the basis of a single review. Peter Dukes, a programme manager at the MRC's London offices, insists that the case is not representative of "the standard to which we work".

There was a happy ending for Roberts — he has since obtained a position at his university that will allow him to set up a lab. Some of the work outlined in the grant has been completed and published in prestigious journals. But, looking back, Roberts recalls how he thought about quitting science when the MRC application was rejected. Bureaucratic failings could have cost Britain a talented scientist.

A better way?

As much as researchers would like to be funded without the hassles of peer review, none would argue in favour of actually distributing public research money in this manner. Funding agencies need to make scientists justify their ideas to make sure that the money goes to the right people. But they also need to minimize the pain involved. How can they avoid some of the incidents mentioned here?

In many cases, US organizations lead the way. In the mid-1990s, the National Science Foundation introduced FastLane, an online application procedure that has won praise for

saving researchers' time, in part because the system automatically picks up problems with formatting issues. Other US agencies are following suit, and eventually all applications to federal funders will be processed through a single website. British researchers have given the thumbs up to a similar website introduced by the country's research councils last year, and Germany's main granting agency, the DFG, has plans for the same sort of approach. Agencies that aren't exploring such possibilities, including some of Japan's government ministries, are beginning to look like the odd ones out.

Other US initiatives have also earned praise, in particular the frequency with which staff attend conferences and talk to scientists. One climate researcher, for example, says that she meets officials from US funding organizations more often than she does their UK counterparts, even though she is based in Britain and receives funding from British agencies.

 **142**
The number of pages, from six different documents, that applicants to the European Union Framework Programme are advised to read before they begin the application process.

"We manage to have only a few days out of the office every month," acknowledges Dukes. He points out that all UK research councils work under a tough cap on the money they can spend on staff activities rather than grants. Instead, Dukes says, the MRC is exploring how internal reforms could help existing staff to get out more. "We want to do better," he says.

But the reforms that generated the most plaudits were those that reduced the number of forms, instruction documents and guidelines that researchers have to plough through. This summer, for example, the US National Institutes of Health (NIH) assessed the impact of a simple change it made six years ago. Applicants who ask for less than \$250,000 in research costs have to describe their project in detail, plus how much funding they require, but the detailed accounting is handled later on at the applicant's research institution.

More than 80% of scientists surveyed about the change say they were satisfied with the policy and about three-quarters of applicants to the NIH's most commonly sought grant, the R01, no longer supply detailed costs in advance. Peer reviewers were similarly enthusiastic, saying that the changes helped them to focus on the pros and cons of the grant itself. Perhaps most importantly, peer reviewers say the process reduced the administrative burden that comes with acting as a referee.

If the NIH system has a failing, say

researchers, it is the length of the application forms. For example, a research outline in an R01 can run to 25 pages. NIH officials say that this is the maximum allowed and that scientists are free to write less. But the psychology of application behaviour seems to dictate that researchers fill all the available space in the hope of winning over reviewers. Simply cutting the length of the form could save time, suggest researchers who have applied for grants elsewhere. In Britain, for example, the equivalent of the R01 involves a form that is a maximum of eight pages long.

When it comes to the grant types on offer, researchers also say that fewer is better. The MRC has recently reduced the number of grant types it offers but made them more flexible, so that researchers can apply for different durations of support, for example. Dukes says the plans are generating positive feedback. Contrast that with the situation in Japan, where the number of different grant types rose from 28 in 2004 to 37 this year.



"You name it, we have heard it."

Walter Schaffer, acting director of the NIH's Office of Extramural Programs, describes the range of complaints he gets from scientists.

In many cases, successful reforms follow a simple recipe: scientists like it when agencies are open, accessible and make the effort to understand them. As one university official who helps researchers with grants notes, it makes a big difference simply to provide a telephone helpdesk that is manned by people (see 'The inside track', right). Scientists also like it when agency staff come to conferences and visit labs, or convene focus groups to ask their opinion on agency practices. Most of all, they like it when agencies actually act on the feedback they are given. Grant applications may be the bane of many scientists' working lives. But they need not be, as research agencies who listen and respond are finding. ■

Jim Giles is a senior reporter for *Nature* in London. Additional reporting by Andreas von Bubnoff in Washington, Ichiko Fuyuno in Tokyo, Tamara Gruener and Valeska Stephan in Munich, and Declan Butler in Paris.



L. WICHSE

The inside track

Meet Karen Bergeron, an administrator at the University of Washington in Seattle with over 20 years' grants experience. She thinks funding applications are fun.

Are US grant applications getting more convoluted?

Sponsoring agencies are having a harder time getting money from Congress, so they are asking scientists to justify their work. You really have to come out and say what the intellectual merit of your project is and what the broader applications are. They want that right at the beginning of the proposal. It's not enough just to describe what you plan to do. I realize a lot of scientists find this insulting — just like they always hate doing progress reports.

One thing that has got easier is that sponsors are a lot more flexible now. In the old days, you had to go back and get approval if you wanted to make changes to the way you were going to spend the money. It's harder now to get the money in the first place, but when you do, you have more control.

Which US agency takes the longest to apply to?

I think it's the National Institutes of Health (NIH), because there are always last-minute things with the proposal — such as getting the pagination and the table just right. It's pretty anal. You can't have staples or some kinds of binding, but you can have binder clips or rubber bands. And your font has to be either Arial or Helvetica at a point size of ten or more.

It's got to the point where people are calling me the font police. I was training someone the other day and he teased me about running an NIH boot-camp, dressing in camouflage and saying: "You too can get an NIH grant proposal out".

The National Science Foundation (NSF) was the first to introduce electronic submission, in the form of FastLane. Does it help?

Yes. Before FastLane, it was not uncommon for proposals to be returned out of hand because of mistakes. One scientist used proportional spacing — he thought it met the requirement. But the NSF took the ruler out and found that there were too many characters to an inch. FastLane does a lot of that checking for you.

And the one thing I love about the NSF is that it has a help desk, and it's always manned by people. You never get a machine — and those people are really smart. There is no substitute for human intelligence.

Most scientists find grant applications infuriating. Do you like your job?

I just think that grants are so much fun. Maybe because things are always changing. You are always working with new sponsors, and your clientele is always different: people come and people go. I love it, it's intellectually satisfying and it's an adrenaline rush when you find out, the second or third time you apply, that this research is getting funded. And even though I am not going to Africa to work with elephants, it's nice to know that is happening.

Emma Maris

Megacity, mega mess...

The creaking infrastructure of Indonesia's capital is overwhelmed by people, vehicles and pollution. As urbanization gathers pace across the developing world, **Jessica Marshall** visits Jakarta to witness its stomach-churning consequences.



On the streets of this city, you can pick your poison. Clouds of black and blue-white smoke billow from the exhaust pipes of buses and motorcycles. Thirteen rivers flow northwards to Jakarta Bay, each a slurry of human waste and garbage. Scavengers pick through the city's rubbish looking for recyclable plastic and cardboard. What they can't sell, they burn — batteries, rubber shoes and all. Rising smoke from burning garbage wafts between the city's skyscrapers.

Pollution in Indonesia's capital, Jakarta, is easy to see, and the causes are not hard to pinpoint. But the effects on its inhabitants' well-being are harder to quantify. Official data are scant, studies of environmental health are few, and those worst affected — the urban poor — are the least likely to be included in city records. Environmental scientists say that much could be done to improve living conditions for those most at risk from pollution. But without a stronger emphasis on research into urban public health, and the political will to act on its findings, experts are pessimistic about making rapid progress. "In the near future, there will be more environmental problems," says James Woodcock, a wastewater consultant to the World Bank who has

lived in Jakarta for more than two decades.

With a population of about 12 million — rising to 21 million if you include the wider conurbation of surrounding towns — Jakarta is already one of the world's largest urban areas. The population of this 'megacity' is predicted to grow by a third in the next decade, part of a global trend towards urbanization. By 2007, the balance of the world's population will tip to give a majority residing in towns and cities¹. Most of the fastest-growing cities are in developing countries (see Chart, opposite). So Jakarta may provide a pointer to a future in which urban pollution becomes a main player in the disease burden of the developing world. "The urban physical environment is going to represent a major health threat," says David Vlahov, an epidemiologist at Columbia University in New York, and president of the International Society for Urban Health.

Gridlock

In Jakarta, air quality is already at crisis point. To get an overview, I meet Budi Haryanto in his wife's office building on a Friday evening in late July. Haryanto, a professor of public health at the University of Indonesia, is waiting for the worst of the traffic to subside before driving home to a Jakarta suburb, a journey of

23 kilometres that can take almost two hours. Some two million people commute into the city each day. From a ninth floor window, Haryanto and I look down on a highway on which stalled head- and tail-lights extend as far as we can see in either direction.

"Jakarta is getting worse," says Haryanto. Traffic is responsible for more than 70% of the nitrogen oxides and particulate matter emitted into the city's air². Haryanto is frustrated that the government is not doing more to monitor and reduce the thick, nostril-burning smog, or to characterize its effects on health. "The Ministry of Health doesn't care," he laments, noting that it is dissolving its subdirectorate dealing with air pollution.

The limited available data paint an ugly picture. Respiratory inflammation accounts for 12.6% of deaths in Jakarta, twice the proportion in the rest of the country³. And estimates based on reported pollution levels attribute more than a million asthma attacks and several thousand premature deaths per year in the city to airborne soot and other particles².

Aside from the sheer volume of traffic, the main problems are poor fuel quality, and a failure to equip vehicles with emissions-control technologies such as catalytic converters. There have been some small steps forward:

SOURCE: REF. 1

before 2001, many vehicles in Jakarta used leaded fuel. At that time, about 35% of Jakartan elementary school students had levels of lead in their blood above the World Health Organization (WHO) safety guideline of 10 micrograms per decilitre⁴. This has now dropped to less than 3%, according to Haryanto's preliminary measurements. But he is concerned that the compound that replaced lead creates emissions of benzene, a known carcinogen. "I suggested to the government that they monitor benzene in the air," says Haryanto. "But they said: 'No funding'."

Although Jakarta's horrendous air quality is evident from a high-rise window, experiencing the city's problems with water pollution and solid waste requires an excursion to street level. Kampung Kandang, a north Jakartan slum, faces a river and backs on to a swamp. I stand on the riverbank, watching the eerily still water slip by. A film of grease coats the surface, broken by plastic bags and other detritus. To avoid paying for garbage collection — which is intermittent, anyway — people drop their rubbish in the river. Downstream, a barrage of trash has collected on an obstacle. The sulphurous smell is overpowering. Next to me, a man flings a wokful of oil into the water.

Dirty old town

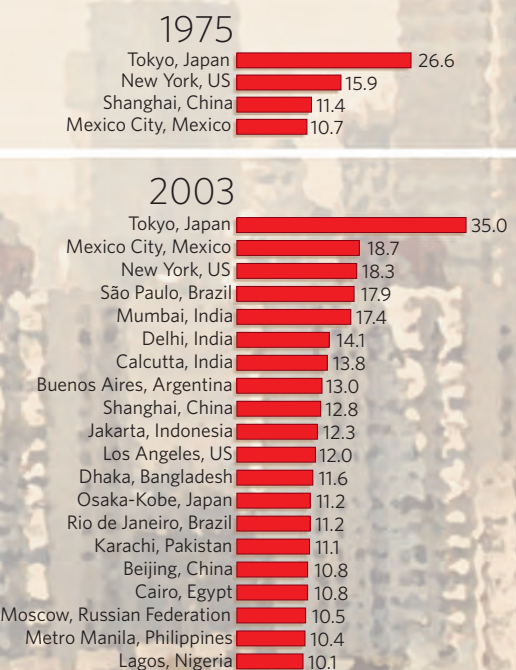
Kampung Kandang is typical of the illegal squatter settlements that line rivers and railway tracks throughout Jakarta, or sit tucked beneath the city's flyovers. It is a microcosm of the city's problems with water, sewage and solid waste. To the rear of the settlement, I watch a chicken in the swamp, scratching on an undulating surface of garbage, oblivious that it isn't on solid ground. The communal water tap opens into a bucket that hangs right above the swamp water that residents use as a latrine. Nearby, an elderly woman wades in the water, collecting swamp plants to sell for wicker.

The public toilet in Kampung Kandang costs up to US\$0.10 to use — no small sum for a family living on about US\$2.50 a day. "So people just do it everywhere," says community leader Miftahul Falah. Water pressure from the tap is low, Falah adds, so the villagers rely on water vendors, who sell 60 litres of water for about US\$0.20 — several times what wealthy Jakartans pay for water from a utility company.

Even for legal residents, supplies are limited. Piped water reaches less than 60% of Jakartans, and is safe for drinking only after being boiled. About half of the supply is lost because of illegal connections and leaks. Water shortages

MEGACITIES WITH MORE THAN 10 MILLION PEOPLE

(population in millions)



have led many residents to tap into ground-water beneath the city. As a result, salt water is seeping into the aquifer, and subsidence has caused parts of the city to sink by a metre or so over the past decade.

Garbage-clogged waterways and the fact that about 40% of Jakarta now lies below sea level conspire to cause annual floods. These hit the poor, low-lying north of the city particularly hard, bringing a litany of health problems. "If the flood lasts a long time, maybe three days," says Falah, "people start to get sick with diarrhoea and rashes."

Less than 3% of the 1.3 million cubic metres of sewage generated each day in greater Jakarta reaches a treatment plant. More than a million septic tanks are buried beneath the city, and these have contaminated most of the city's wells



Scavengers scale the massive landfill at Bantar Gebang seeking things they can recycle and sell.

with faecal coliform bacteria. What's more, truck drivers hired to pump the tanks often dump their loads, untreated, into waterways.

Solid-waste management is similarly chaotic. The city's Bantar Gebang landfill is a case in point — soil is applied only every few weeks and leachate is inadequately treated, says Widhi Handoko, an instructor in solid-waste management at the Ministry of Public Works. An army of 6,000 scavengers works the mountains of garbage. Like post-apocalyptic sherpas, clad in rubber boots and with wicker baskets strapped to their backs, they travel in the wake of bulldozers, plucking recyclables from the stinking heap.

Heaps of trouble

Although Bantar Gebang is nearing the end of its 20-year design lifetime, its representatives say that there is no option but to keep it open while the city seeks alternatives. A private company has developed land for a new waste-disposal site, but local residents have protested loudly. The municipal government recently announced it will build four incinerators. But this is an expensive option, and may cause other environmental and health hazards.

Many people in Jakarta's poor neighbourhoods say their health is fine, despite the filth that surrounds them. But experts believe that poor sanitation is a serious health issue. Ministry of Health records show gastroenteritis is by far the most frequent disease diagnosis at local clinics and hospitals. The incidence of dengue fever has also exploded in recent years. "It is not normally an urban health issue," says Jan Speets, an adviser with the WHO in Jakarta. But flooding and the piles of rubbish throughout the city have created breeding opportunities for the mosquitoes that spread the disease.

Experts in public health urge more and better research to quantify the health problems

J. MARSHALL

caused by poor sanitation and waste management. "There are no real studies available to reveal what's going on in the city," complains Jaap van Dissel, an infectious-disease specialist at the Leiden University Medical Center in the Netherlands. His recent investigation of the food- and water-borne diseases typhoid and paratyphoid in east Jakarta found that doctors over-diagnose the former by up to ten-fold because blood cultures that confirm the infection are not normally done⁵. This illustrates the need to improve clinical diagnoses before attempting potentially expensive campaigns to address problems with public health, says van Dissel: "It's important to know your enemies before you start shooting."

Many of Jakarta's problems are shared by other megacities in the developing world. Most have large illegal shanty towns, and face similar issues with pollution and waste management. For instance, recent flooding in and around Mumbai in India, attributed in part to clogged drainage throughout the city, killed more than a thousand people, and brought water-borne diseases in its wake.

Scrubbing up

Some developing-world megacities have taken steps to clean themselves up. Mexico City's appalling smog is now beginning to clear thanks to the introduction of catalytic converters and improvements in fuel quality⁶. And the Indian capital of New Delhi is experiencing similar gains after converting its public transport to run on compressed natural gas.

So what are the chances of Jakarta following suit? Experts say that solving the city's problems with environmental health will require genuine political commitment to pay for research and monitoring to characterize the problems, and spending on the infrastructure needed to solve them. Given a legacy of official corruption, and the continuing hangover from the Asian economic crisis of 1997, the obstacles are formidable — public spending on infrastructure is running at 80% less than during the heady days of the mid-1990s, when Asia's economy was booming⁷.

So far, politicians seem more interested in sweeping pollution under the carpet, rather than tackling the problems it causes head-on. After the WHO labelled Jakarta the world's third most polluted metropolis in the early 1990s, air-quality monitoring equipment was moved to residential areas with lower levels of pollution.

Ritola Tasmaya, secretary to the governor of

"A film of grease coats the river's surface, broken by plastic bags and other detritus. To avoid paying for garbage collection people drop their rubbish in the river. The sulphurous smell is overpowering."



Kampung Kandang's communal tap opens into a bucket by the trash-choked swamp that doubles as a toilet.

Jakarta, defends the municipal government's record, pointing to developments such as a recently built busway, which will later incorporate new buses running on compressed natural gas. Tasmaya blames continuing problems with environmental health on insufficient budgets and limits to the city government's authority — rivers, he notes, remain the responsibility of the national government. "Jakarta as a capital city needs special support from the central government," Tasmaya concludes. "The infrastructure must be good enough so that people who come here for business, tourism and investment can be served."

Foreign specialists say that significant progress could be made if existing environmental regulations were properly enforced. "It's very difficult for a government that's known to be corrupt to enforce laws," says Woodcock. But the good news is that, after years of dictatorship and corruption, Indonesia is slowly becoming more democratic. Last year, the country gained its first directly elected president, Susilo Bambang Yudhoyono. And 2007 will see the first direct election for the governor of Jakarta.

For now, many of the city's residents have

more immediate priorities than reducing pollution. "Income is still low," says Basah Her-nowo, director of settlements and housing at the National Development Planning Agency, an arm of the central government. "People do not care about environmental quality. They are still thinking about their stomachs." But problems such as flooding and waste mismanagement are getting so bad that people are beginning to call for change. As democracy takes root, environmental health may slowly move up the list of political priorities. "In the end," Woodcock says, "I feel optimistic that there will be progress."

Jessica Marshall is a science writer currently travelling in Asia.

1. *World Urbanization Prospects: The 2003 Revision* (United Nations Department of Economic and Social Affairs, New York, 2004).
2. *Integrated Vehicle Emission Reduction Strategy for Greater Jakarta, Indonesia* (Asian Development Bank, Jakarta, 2002).
3. *Indonesia Environment Monitor 2003* (World Bank, Jakarta, 2003).
4. Albalak, R. et al. *Sci. Total Environ.* **301**, 75–85 (2003).
5. Vollaard, A. M. et al. *J. Am. Med. Assoc.* **291**, 2607–2615 (2004).
6. Molina, M. J. & Molina, L. T. *J. Air Waste Manage. Assoc.* **54**, 644–680 (2004).
7. *Averting an Infrastructure Crisis: A Framework for Policy and Action* (World Bank, Jakarta, 2004).

BUSINESS

Defence group aims for take-off

A share offering beckons this autumn for QinetiQ, a business that is being built out of some of Britain's most famous military research laboratories. **Andrea Chipman** reports.

The low-rise, glass-and-steel research campus that houses QinetiQ in Farnborough, south-west of London, betrays few signs of the company's historic pedigree. It was on this site that the then Royal Aircraft Establishment (RAE) developed many military aircraft that flew in the Second World War, and contributed to the design of Concorde.

But now the company — whose name is pronounced 'kinetic' — is part of a rapid transformation that is converting the former defence labs into a unique, research-led defence company that is making money and will soon stand on its own two feet.

A stockmarket offering — which could take place as soon as November — would be the culmination of a 14-year process that has seen parts of the former labs converted into a dynamic defence contractor and consultancy firm, which draws on research in fields such as materials science and optics to strengthen its competitive position.

QinetiQ was founded in 2001 and is currently a public-private hybrid, with just over half of its shares held by the British Ministry of Defence, one-third by the Washington-based Carlyle Group, and the rest by management and staff. It has reduced its number of sites by half without any net loss of staff, the company says, and most of its 9,000 UK employees now work at its offices at Farnborough and in Malvern, Worcestershire.

And although research will remain central to the company's identity, it is already clear that other activities — primarily military contracting and consulting — will constitute the largest part of its business. Contracts with the Ministry of Defence still account for more than half of its annual sales, but the company has also built a strong presence in the United States by procuring several established contractors to the Pentagon and the Department of Homeland Security over the past year.

"The original plan was to commercialize the defence technology," says one person close to the company's thinking. "But that has been hard to do." QinetiQ will continue to carry out research, but this will increasingly be outweighed by more general defence contracting.

"At the moment, QinetiQ does have a privileged position," says Clive Forestier-Walker, a defence and aerospace analyst at London-



Big business: expertise in aviation provided part of the basis for QinetiQ's progress towards a share offer.

based Numis Securities, citing the company's work on research and development for the Ministry of Defence and as a consultant for defence contractors such as BAE Systems.

"The general view is that QinetiQ will be a highly rated company because of the sectors it is in," says another London-based analyst, who predicts that the company's defence and security interests will attract investors.

But QinetiQ is still looking for commercial uses for military technology. The materials-science unit at Farnborough, for instance, is looking at how stealth materials that were developed for use in military aircraft can find other applications. "Historically we've concentrated on defence projects, but this puts us in a strong position to find commercial work," says Chris Lawrence, a physicist who heads the company's smart-materials group.

One project at the unit's impressively equipped Farnborough laboratory is seeking to combine expertise in optics, infrared,

microwaves, acoustics, visible light and magnetism to develop devices that can be used in counterfeit deterrence and inventory control.

Another potential application of stealth materials is a thin-film material that uses microwave-absorber technology and could be used with radio frequency identification (RFID) tags. RFID can help companies secure and track retail merchandise, and is a growing market that could be worth more than US\$7 billion worldwide by 2008, according to the independent technology consultant IDTechEx.

QinetiQ is looking at ways of using the technology in 'wallpapers' that absorb stray mobile-phone signals, and prevent the scatter of radiowaves when tags are being scanned, Lawrence says. "Our knowledge of stealth technology has enabled us to think of various ways of either making something very eye-catching, or else hiding a message," he explains.

Business opportunities have already flowed

QINETIQ

from the company's activities in nanotechnology. In 2000, QinetiQ's predecessor organization teamed up with Australian biotechnology firm pSivida to form a joint venture to develop BioSilicon, a form of nanostructured porous silicon with medical applications, including controlled drug delivery and brachytherapy, a localized treatment of cancer.

Meanwhile, its subsidiary QinetiQ Nanomaterials is marketing its proprietary process for producing bulk nanopowders from a range of pure metals and other materials, under the brand name Tesima.

Taking shape

QinetiQ's progress has been overseen by John Chisholm, an engineer and former founder and chief executive of CAP Scientific, a London-based software house. Chisholm was asked by the Ministry of Defence to supervise the merger in 1991 of several military research laboratories, including the RAE and the Royal Signals and Radar Establishment at Malvern, into the Defence Research Agency, which then became the Defence Evaluation and Research Agency, or DERA, in 1995.

In 2001, the government held onto the parts of the DERA labs that did the most classified work, and formed QinetiQ to run operations with commercial potential. The next year, it selected the Carlyle Group as a partner to take the new company into private ownership, and sold it a third of the shares for £150 million (US\$275).

Carlyle is a private investment company that has strong ties to the US national security establishment — former president George Bush used to serve on the board — and its involvement has clearly opened doors for QinetiQ as it has expanded aggressively into the United States, buying four defence contractors over the past year. For good measure, it just bought a 90% share of a Belgian space company, Verhaert.

With its fingers in a number of potentially profitable pies, QinetiQ now has to balance its various priorities: government contracts, private commercial contracts, consulting work and innovative research. Its multiple roles can mean that the company is at the same time advising BAE Systems and providing services for the Ministry of Defence and US defence and security agencies.

QinetiQ's unique position is seen by some observers as a strength, and financial analysts have predicted that the share offering will value the company at more than £1 billion. Foreigners can forget about swooping for control of the company, however; the Ministry of Defence will retain a special share that will prevent it falling under ownership of which it doesn't approve. ■

IN BRIEF

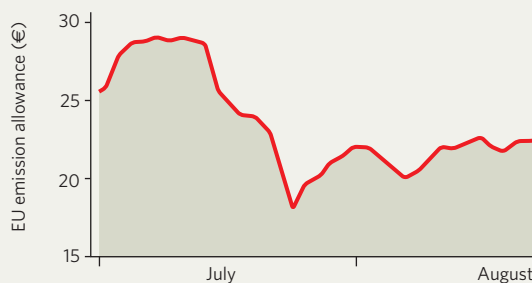
RNAi ON TARGET Drug maker Novartis has bought a major stake in Alnylam Pharmaceuticals — one of a clutch of companies that are pioneering the use of RNA interference in medicine. The biotechnology company, which is based in Cambridge, Massachusetts, will initially receive \$57 million from Novartis for about 20% of its stock, and the two companies will pursue a research collaboration that could provide up to \$650 million more to support Alnylam if its products achieve commercial success. RNA interference uses snippets of recombinant DNA to switch off disease-causing genes.

PCR SPAT BROADENS OUT Applied Biosystems, a manufacturer of polymerase chain reaction (PCR) machines for genetic sequencing, says it has won a court injunction to prohibit a rival from selling or maintaining similar equipment. The injunction from a court in New Haven, Connecticut, stops MJ Research and its parent, Bio-Rad, from making, selling or producing thermal cycler products for PCR machines. California-based Bio-Rad issued a statement saying that it thought it had agreed otherwise with Applied Biosystems, and that it is "dismayed" at the injunction.

CASH INJECTION GlaxoSmithKline is to buy ID Biomedical, the Canadian vaccine manufacturer, for \$1.4 billion. Both companies supply flu vaccines, and the deal seems to reflect growing commercial interest in vaccine manufacture. In a related development last week, Chiron's independent directors rejected a \$4.5 billion bid from Novartis to buy full control of the California-based vaccine maker, in which it already holds a large minority stake.

MARKET WATCH

EUROPEAN CARBON INDEX



After crashing drastically in July, the costs of options to emit carbon dioxide in Europe have been steadily rising again through August and early September.

Analysts blame the crash on a rush to the exits by speculators who had been investing in the options — and warn that further volatility could lie ahead, as more nations and investors join the embryonic market for the options.

At the European Energy Exchange (EEX) in Leipzig, Germany — one of five such exchanges now operating across the continent — the price of an allowance to emit one extra ton of CO₂ peaked at €29 (US\$36) in early July, before falling back to €18. Last week, the allowances were back up to €25, with some brokers predicting that they could soon rise beyond €30.

But investors should be ready for

future surprises, says Gabriele Rahn, a spokeswoman on energy trading at energy company Vattenfall, based in Hamburg, Germany. "The European-wide trade volume has come along nicely," she adds.

More than one million emission allowances per day are currently being traded across the five exchanges, Rahn says. But some countries have yet to start participating in the European Union's emission-trading scheme, which was set up in January. Once the latecomers, including Italy and Poland, join in, both trade volume and market volatility could increase again, she warns.

July's plunge caught analysts by surprise. According to one London-based broker, who declined to be identified, it came about after investors who had been buying the allowances in search of a short-term profit decided to get out.

Quirin Schiermeier

SOURCE: EUROPEAN ENERGY EXCHANGE

Re-wilding: introductions could reduce biodiversity

SIR — The Commentary article by Josh Donlan and colleagues (“Re-wilding North America” *Nature* **436**, 913–914; 2005) argues for the introduction of old-world mammals to North America, on the grounds that these are proxies for megafauna that lived there at the end of the Pleistocene. This perspective overlooks environmental changes that have occurred during the intervening millennia, and that have produced qualitatively different communities in new ecological equilibria.

For example, global climate change since the Pleistocene extinctions makes the restoration of vanished ecosystems through large-mammal introduction quite unlikely. Such environmental change also increases the risk that introduced species might respond in unexpected ways. Indeed, one of the reintroductions proposed by Donlan and colleagues — the camel — was previously attempted during the nineteenth century (T. L. Connelly, *Southwestern Historical Quarterly* **69**, 442–462; 1966). At that time, camels seemed an ideal beast of burden for use in the North American arid regions, but American environments ultimately proved inhospitable. By 1900, both public and private introduction programmes had failed, and the remaining feral camels had died.

The inherent unpredictability associated with disturbing ecosystems means that, while some introductions might follow the camel's fate, others might prove all too capable of adapting to contemporary North American environments, potentially at the expense of other species of conservation value. Indeed, introduced species are now a major cause of biodiversity losses worldwide.

Christopher Irwin Smith

Department of Biological Sciences,
University of Idaho, Moscow, Idaho 83844, USA

Re-wilding: other projects help carnivores stay wild

SIR — In their plea for bringing Pleistocene wildlife to the New World (“Re-wilding North America” *Nature* **436**, 913–914; 2005), Josh Donlan and colleagues do not discuss successful efforts to ensure long-term survival of large carnivores in Africa and Asia.

In Namibia, the Cheetah Conservation Fund has developed programmes to foster acceptance of this predator, by providing farmland-owners with educational material and encouraging them to take pride in cheetah presence. The number of cheetahs removed has dropped from 19 to 2.1 per farm per year since 1991. Ranchers enrolled in the programme can also export beef, certified ‘cheetah friendly’, to the European Union —

making cheetah protection both ecologically possible and economically profitable.

In Kenya, a study shows that *bomas*, traditional corrals with thick walls and internal rooms, are effective at reducing the amount of livestock killed by lions (M. O. Ogada *et al. Conserv. Biol.* **17**, 1521–1530; 2003). This finding has the potential to reduce conflict with humans, which is the main threat to lion survival.

In central Asia, the International Snow Leopard Trust and the Snow Leopard Conservancy provide incentives to local herders for protecting snow leopards. These include insurance against damage by carnivores, veterinary care for livestock and income generation from handicrafts. Participating herders in Mongolia, the Kyrgyz Republic and Pakistan agree not to kill snow leopards or their prey, in exchange for access to foreign markets to sell labelled knitwear. The success of this programme owes much to peer pressure, as the whole community loses the bonus if one person violates the contract. In the eastern Kyrgyz Republic, this has led to the first year with no recorded poaching since the collapse of the Soviet Union.

We believe that these diverse pilot schemes will ensure that large carnivores in Africa and Asia have a good chance of persisting in the wild into the next century.

Guillaume Chapron

Laboratoire d'Ecologie CNRS UMR 7625 et
Plateforme Environnement, Ecole Normale
Supérieure, 24 rue Lhomond, 75005 Paris, France
Laboratoire d'Ecologie Animale UMR MA105,
Université d'Angers, France

Malaysia can't thrive while it excludes minority talent

SIR — As a Chinese Malaysian, I sympathize with the sentiments described in your News story “The Valley of Ghosts” (*Nature* **436**, 620–621; 2005) about the Malaysian government's often-denied ‘Malays first’ policy. In particular, university admissions are suspicious. Malay admissions are based on results from exams called *Matrikulasi*, taken at Malay-only pre-university colleges, whereas other students have to take the national Malaysian Higher School Certificate (STPM) exam. Reports of ethnic-minority students with near-perfect STPM results not getting a place at the local university have become the norm, and yet objections are often ignored — the government claims that it's a fair game for all.

Personally, I had no choice but to go overseas to study, and my parents had to spend their entire pension savings on financing my undergraduate degree in Australia. After graduation, most of my Malaysian classmates chose to either stay in Australia or work in Singapore, where fair

competition and equal opportunities give them better job prospects. Before coming to the Netherlands I did my master's degree in Singapore, where I met many Chinese Malaysians in this situation.

Most of us would like to return to Malaysia, but we know that research prospects for minorities are limited. No matter how talented we are, it seems we still have to travel outside our country to seek opportunities.

In Yee Phang

Department of Materials Science and Technology
of Polymers, University of Twente, PO Box
217,7500 AE Enschede, The Netherlands

Harry Potter and the prisoner of presumption

SIR — Jeffrey Craig and colleagues, in Correspondence (“Harry Potter and the recessive allele” *Nature* **436**, 776; 2005), recommend the use of analogies as tools for introducing young people to scientific concepts. Taking their example from J. K. Rowling's stories about the young wizard Harry Potter, they suggest that wizarding is a monogenic trait, with the wizard allele (W) recessive to the muggle allele (M). We believe the assumption that wizarding has a genetic basis to be deterministic and unsupported by available evidence.

Following Craig and colleagues' analogy, Hermione, as a muggle-born witch, must have WM parents. However, as Rowling fans could point out, Hermione's parents were muggle dentists who lack any family history of wizarding. It's true, of course, that chance may not have thrown up a witch or wizard for many generations, or that any who did have magical powers may have kept them secret to avoid a witch hunt.

What about Neville's apparently poor wizarding skills? These cannot be explained by incomplete penetrance, as Craig and colleagues suggest. In incomplete penetrance, individuals either display the trait or not: they do not display an intermediate degree of the trait. Poor wizarding skills might be indicative of variable expressivity of an allele. However, both variable expressivity and incomplete penetrance are associated with dominant alleles. If the wizarding allele were dominant, rather than recessive as suggested, wizarding children such as Hermione could not be born to non-wizarding parents.

Neville's clumsiness may, perhaps, be an individual characteristic unrelated to his potential powers. However, it is not possible, from the evidence presented so far, to conclude that wizarding is a heritable trait.

**Antony N. Dodd, Carlos T. Hotta,
Michael J. Gardner**

Department of Plant Sciences,
University of Cambridge, Downing Street,
Cambridge CB2 3EA, UK

BOOKS & ARTS

2FACE/CORBIS

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Fear of the future

Will scientific innovation bring progress and benefits, or just risks and dangers?

Unersättliche Neugier: Innovation in einer fragilen Zukunft. [Insatiable Curiosity: Innovation in a Fragile Future]

Helga Nowotny

Kulturverlag Kadmos Berlin: 2005. 203 pp.
€19.90

Hubert S. Markl

Helga Nowotny is not only *la grande dame* of science studies in Europe, she is also one of the most savvy and influential people in European research affairs. She chairs the European Research Advisory Board and is a board member of the nascent European Research Council. Reason enough, then, to turn to this book with high hopes (or even, to borrow from the title, “insatiable curiosity”) for her deliberations on science, technology, innovation and the human future.

I must admit though that this half-popular, half-scholarly essay made rather uncomfortable reading for a dyed-in-the-wool natural scientist like me, who finds himself the guinea pig of science sociology studies. A psychoanalyst might say that such resistance is the first sign of trouble with our scientific-rational world-view. Or, as German chemist Justus von Liebig once remarked: “I seldom have a good idea, but if someone else comes up with one, I immediately have a better one!” However, it is useful to see how a highly knowledgeable sociologist of science looks at our science through the lens of her discipline. I assume Nowotny had precisely this in mind: to incite readers from any persuasion to argue emphatically about the issues she raises in this book.

What are these issues? The subtitle says this is a book about innovation and its decisive role for our unpredictable future (I wonder what is meant by “fragile” future — has it ever seemed anything else?). Nowotny tells us a great deal about the sources of scientific and technological innovation and its increasing influence on economic competitiveness in a globally accessible world. Solving problems and serving desires, and thus creating new problems to be solved, with respect to energy supplies and world climate, resource depletion and waste accumulation, water and food availability, pandemics, the flood of global media and rising social unrest. All this is argued persuasively, although not always in a novel way, and there is a sense of anxious urgency, like that of a rodeo rider clinging to the back of a bucking bronco.

The book contains some interesting historical vignettes and clear-sighted comparisons between biological and cultural innovations. They strongly emphasize symbol technologies, although strangely the book neglects the human achievements that drove the most innovation: tool-making and language. In fact the whole exercise seems somewhat mistitled: the book seems profoundly ambivalent to innovation. Of course, like other texts from the sociology of science, it is not so much a book on science as on writing on writing on science, far enough removed from the research enterprise to take the sometimes rather supercilious attitude occasionally found in research on research.

Above all, the book never fails to chastise the “hubris of believing in progress” — that deeply

flawed illusion of the past centuries — while passing over the doubling or tripling of life expectation, the abolition of regular mass starvation in many formerly stricken countries, the conquering of diseases such as smallpox and poliomyelitis in large parts of the world, the disappearance of many horrendous superstitions, and so on. These achievements are presumably not even worthy of notice, as all of this and much more is taken for granted. This is not progress, but entitlement, according to those critical of progress, although strangely enough these are not goods received from caring gods, but from that progress-blinded sci-tech civilization. The fence between pro- and anti-science, and pro- and anti-innovation, seems to be firmly straddled here — maybe not the most pleasant place from which to dwell on thorny issues. Is it not difficult to both have one's cake and discard it as garbage?

Nowotny makes the point that our future is wide open to risks and dangers of our own making, as we try to steer between 6 billion and 9 billion humans through the uncharted waters of their unknown destiny. She emphasizes correctly the increased volatility of too many of the foundations of our wheelings and dealings. But I wonder whether the future was any less unpredictable for those ancient women and men, scared by the vicissitudes of only too certain failed harvests, plagues or threats from fellow beings. Such scenarios cannot have been less menacing than those of our innovation-bound societies. Of course, if you include the religious promise of eternal life after death, life expectancy wasn't quite as bad back then, as historian

Arthur Imhof has remarked. But when humanists belittle the progress made in the past few centuries, I doubt that they would have us regress to such pre-Enlightenment conditions.

This book seems to emanate a feeling of suffering from modernity, while emphasizing that innovation will be the inevitable hallmark of modernity (or rather, postmodernity, as the dark alley ahead of modernity will always have to be called). As the Roman historian Titus Livius succinctly put it more than 2,000 years ago, "*Nec vitia nostra nec remedia pati possumus*" ("We can endure neither our vices nor their remedies"), which shows that this ambivalence is not so recent.

There are a few minor points to be raised. First, it seems regrettable that this essay from a leading European science-policy figure has not been published in English. Maybe this is because the mixture of socio-scholarly, doubt-

ridden, intellectual *Zeitgeist* and *Menschenbild* worries is only too German? It is to be hoped, however, that this is not the *Menschenbild* exemplified by the art of Patricia Piccinini on the cover of this book, which depicts a young family of pig-like humans or humanized pigs! If there is to be an English edition, hopefully minor errors, such as the claim that prokaryotes evolved from eukaryotes (it was the other way round), or the figures for derivative financial markets, which mix up US trillions and the German *Trillionen*, can be corrected.

Such minor quibbles aside, this is a very readable book. It is thought provoking, but also incited me to disagree with some of its doom-laden messages. Insatiable curiosity? Let's hope so, under the challenging demands of unending necessity. ■

Hubert S. Markl is in the Department of Biology, University of Konstanz, 78457 Konstanz, Germany.

his chronicle a now well known statistical forecast: "Red sky at night is a shepherd's delight." Montesquieu, in contrast, studied the taste buds of sheep's tongues and their blood circulation at various temperatures, and concluded that northern people were bold and not very devious or sensitive compared with those from lower latitudes!

This kind of thinking and eccentric collection of data continued until the early twentieth century when geographers, historians and anthropologists pointed out that societies evolve as much through organization and religion, for example, as through climate.

Boia brings a topical dimension to his perspective when he emphasizes the relation between the way societies have dealt with climatic events and with natural disasters. In their reactions to the sudden loss of life and disruption associated with the latter, most societies have sought religious explanation. The Bible and other early writings focused on whirlwinds, fire, earthquakes, floods and droughts. They also revealed how various kinds of 'divine intervention' have helped or hindered the hazard, depending on the point of view: the Japanese, for example, are grateful for the 'kamikaze' typhoon that saved them from Genghis Khan. The ice age was the last globally significant climate change that humans endured. It was also a natural disaster of huge proportions as the ice retreated some 10,000 years ago. This shaped the landscape of Britain and was probably associated with floods in the Middle East, India and the north-western United States (where the mythical raven god carried people away on its wings).

This extreme form of climate change was feared by religious alarmists to be imminent at the start of the sixteenth century. This led the president of the Toulouse parliament in France to use the famous woodworking skills of the region to build another ark. According to the Michelin guide, these skills later led to the Lagardère media company and Airbus.

In the concluding chapter on climate change, Boia sides with Bjørn Lomborg in suggesting that a 2 °C change in global temperature is

A climate for social change

The Weather in the Imagination

by Lucian Boia

Reaktion: 2005. 224 pp. £14.95

Julian Hunt

It's probably only fair that *Nature* should publicize the views of a historian about meteorology, because in the past it has published influential letters by meteorologists on history. Lewis Fry Richardson demonstrated that differential equations and statistical laws that can successfully model weather systems should also be able to model humans' behaviour and maybe even psychology, from their proclivity for conflict to their appreciation of jazz. This approach not only explains quantitatively how wars did or did not develop, but in 1935 and 1951 predicted future developments (see *Nature* 411, 737; 2001).

Lucian Boia, a historian at the University of Bucharest in Romania, has written a stimulating book, *The Weather in the Imagination*, reviewing the literature on theories of how climate has affected societies, and of how humans may have influenced climate. He concludes with a personal, if not entirely accurate, account of the science of human-induced climate change, and debates the current policy options. He reveals his methodological bias, however, when he implies that Newton's work on predicting the movement of planets is a rather simple matter compared with studying the complexity of history. (Richardson, in contrast, had noted that, like the natural world, societies can in some respects have simple mathematical descriptions, for example in the way that armaments can grow exponentially before a war, and that the frequency of conflicts tends to follow a Poisson distribution.)

The big historical question in this book —

the extent to which national characteristics are determined by the weather and climate — has been addressed in fascinatingly different ways by Hippocrates, historian Edward Gibbon, French philosopher Montesquieu, the Arab historian Ibn Khaldun, and more recently by Hubert Lamb and Emmanuel Le Roy Ladurie. Much of the evidence is anecdotal and rather surprising. The Greeks thought that the cold weather made the British not only aggressive but sexually promiscuous — obviously the directors of recent reality TV shows could have saved themselves a lot of money by staying in Britain, if only they had read their Hippocrates. Gibbon extended this climate hypothesis by arguing that people from northern Europe had also been influenced by the way they modified their climate through deforestation and agriculture. Some writers included real weather observations: the Venerable Bede in the eighth century included in

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

More sex please, we're British: did the climate, painted here by Constable, make people promiscuous?

of no great concern. A conference of meteorologists held in Exeter, UK, in February (www.stabilisation2005.com) disagreed, concluding that such an outcome would be unacceptable to most societies if they had a choice in the matter. Boia himself points out that such a rise in temperature will lead to the loss of mountain glaciers and the destruction of millions of species of plants and animals. He makes a technical error, however, in stating that the reduction of emissions needed to mitigate this rise can be provided solely by the more economical and efficient use of

transport. These steps, although necessary, are insufficient: the excessive use of energy for heating and cooling buildings, which in most countries gives rise to 50% of emissions, also needs to be curbed.

Lomborg is correct that the local environment and living standards are improving for many people. But some climate models would see people reduced to communities perched on hilltops in a depleted natural environment. Some of us might be living on the water in the next generation of ark, indeed the Netherlands is planning to develop floating houses for use

in the most flood-prone areas of the country.

An alternative reading of history might conclude that human societies can rise to extraordinary challenges, as the 3,000-year-old society of China surely demonstrates, and should therefore be able to avert the climatic consequences of our actions and prevent their worst effects. Unfortunately that is not the message of this important book. ■

Julian Hunt is in the Centre for Polar Observation and Modelling, Department of Space and Climate Physics, University College London, Gower Street, London WC1E 6BT, UK.

Early views of viewing

The Moving Tablet of the Eye: The Origins of Modern Eye Movement Research

by Nicholas J. Wade & Benjamin W. Tatler

Oxford University Press: 2005. 312 pp.

£75, \$145 (hbk); £29,95 (pbk)

John M. Findlay

"Why is it to those who are very drunk everything appears to revolve?" Nicholas Wade and Benjamin Tatler use this question, once posed by Aristotle, to illustrate visual vertigo, the illusory motion of the visual world. A further challenging question is why, during our more sober moments, do we not have such experiences, and instead perceive a stable world, despite the continual changes of retinal stimulation resulting from the motion of our eyes.

Since at least the time of Johannes Kepler, some 400 years ago, research in vision has been based on the image-forming eye. The apparent close correspondence between our perceptual experience and our retinal image has generally led to the second question being answered in terms of the 'suppression' of, and 'compensation' for, the changes caused by the eyes' motion. Yet these terms, particularly the latter, presuppose some form of 'inner screen' representation. As the authors point out, the intuitively appealing idea of a detailed mental copy of our visual environment is increasingly recognized as a 'grand illusion'. Some researchers, notably J. Kevin O'Regan and Alva Noë, even reject any form of internalized visual representation. In line with this reappraisal there has been more appreciation of the importance of the mobility of the eyes in visual science; Wade and Tatler at one point contend that "eye movements lie at the heart of contemporary studies of vision".

The Moving Tablet of the Eye is a chronicle of the history, reinterpreted by modern thinking, of studies of seeing actively and of the many contributions to the field from the ancient Greeks to the early twentieth century. One of its idiosyncrasies is the predilection for providing physiognomic rather than geographic information about the scientists involved. The book contains more than sixty miniature

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

We normally perceive a stable world even though our eyes keep moving.

portraits, but biographic information other than dates is provided only for the major players, such as William Porterfield, Charles Wells, Émile Javal and Raymond Dodge.

The first two names highlight the advanced level of intellectual activity in Scotland in the eighteenth and early nineteenth centuries. Porterfield, an Edinburgh physician, certainly appreciated the importance of eye movements. In a publication in 1737 he decried the "Vulgar Error" of assuming that we see everything distinctly, clearly and at the same time. Wells, who was educated in Scotland but worked in London, demonstrated (using extended vestibular stimulation, rather than inebriation) that visual vertigo is linked to eye motion. His use in the 1790s of after-images for this purpose anticipated the major technique used to study eye movements during the following century. Wells successfully defended his position against a challenge from the august figure of Erasmus Darwin, whose integration of speculative science and ponderous poetry is well represented in the book and provided the inspiration for the title.

Another notable piece of historical research by the authors reveals the initial use of the term 'saccade' and the first appreciation of the jerky quality of eye scanning. Javal, a distinguished French ophthalmologist, is now often

per line of text is unchanged, regardless of the viewing distance. However, he didn't get round to publishing until 1892. In fact, no less a figure than Ewald Hering carried out a similar experiment, correlated the resultant "dull clapping" sounds with after-image movements, and published the finding, also in 1879.

This book is extensive and thorough but not exhaustive. The history largely stops in the early years of the twentieth century, so the near absence of any oculomotor neuroscience is justifiable — although the fascinating neurological condition of 'psychic paralysis of gaze' might have merited a mention. Another surprising omission is the work of E. E. Maddox on the preconditions for vergence movement. The authors have made every effort to integrate the study of looking with that of seeing, but the malign influence of the grand illusion can perhaps still be detected when we read that the saccade-and-fixate strategy "evolved as an adaptation to the demands of a highly mobile eye". But these are minor blemishes on a fascinating work and a splendid scholarly achievement. The book will surely stand as the definitive text on the history of eye-movement research for many years to come. ■

John M. Findlay is in the Department of Psychology, University of Durham, South Road, Durham DH1 3LE, UK.

credited with both, thanks to a careless attribution in Edmund Huey's classic text on reading. The term saccade was first used in an oculomotor context in Javal's writings in 1879, but in a footnote describing an observation by A. Lamare, a co-worker in Javal's laboratory at the Sorbonne. Using a mechano-acoustic transducer, Lamare heard noises corresponding to the discontinuous movements of the eyes during reading, and observed that the number of saccades

T. WATSON/ALAMY

Dirac's hidden geometry

Paul Dirac insisted that his approach to quantum physics was geometric not algebraic. But where is the evidence of this in his pioneering, algebra-rich papers?

Graham Farmelo

Of the few authentic visionaries modern science has known, Paul Dirac was the most inscrutable. He was a man of legendary quietness and privacy. His peers were bemused by his strikingly unusual way of looking at the world, while also amazed by the fecundity of his methods. Beginning in 1925, Dirac spent eight years developing quantum mechanics in a series of elegant papers that repeatedly took theoretical physicists by surprise. What, they wondered, lay behind Dirac's unique methods?

But colleagues seeking clues about the provenance of Dirac's ideas found none in his papers. He began each one with a deceptively straightforward, equationless introduction, before steam hammering his way through the mathematics, giving no quarter either to the faint hearted or to pedants.

Few of his papers contained a diagram and none offered any solace to readers trying to visualize what was represented by the blizzard of abstractions. Dirac would respond to requests to express his quantum reasoning in words or pictures by impassively shaking his head. So it is hardly surprising that Dirac's peers perceived him to be, above all, a brilliant algebraist — extremely adept at manipulating abstract symbols, but uninterested in visualization.

Yet Dirac insisted that he was not primarily an algebraic theoretician. This insistence first emerged almost 40 years after his first quantum papers were published, during an interview with the historian of science Thomas Kuhn in the spring of 1963. Dirac declared to Kuhn that his approach was fundamentally geometrical and that he was “just no good” at doing “masses of algebraic calculations without picturing what the equation means”.

Robert Oppenheimer, Dirac's close friend and admirer, was incredulous when he heard this. Oppenheimer had hardly ever seen Dirac draw a diagram but had always been awed by his algebraic skill. No, Oppenheimer assured Kuhn, Dirac was “principally algebraic”.

Indeed, Oppenheimer went further and commented that Dirac's approach to physics was intuitive, like the mumbling savant Niels Bohr, although the two had completely different ways of communicating. “Bohr

regarded mathematics as Dirac regards words, namely as a way to make himself intelligible to other people, which he hardly needs,” he said.

In a scientific memoir nine years later, Dirac divided mathematicians cleanly into algebraists and geometers (the field of analysis apparently did not count). He had observed that mathematicians with a European training tended to be interested in geometry, following the school of the ancient Greeks, whereas those from an Asian background preferred algebra, which

reference in Werner Heisenberg's pioneering paper to the existence of pairs of quantities with the strange property that $A \times B$ is not always equal to $B \times A$. Darrigol points out that this type of relationship occurs in the elementary theorems of basic projective geometry, which Dirac had certainly studied. So Dirac would have been ready to capitalize on the analogy and to develop a geometric picture of Heisenberg's theory. An intriguing explanation, but is it correct?

Dirac had a perfect opportunity to clarify the point at a talk he gave in the autumn of

1972 at Boston University. The university's philosophy department had brilliantly taken the initiative of inviting Dirac to explain for the first time how projective geometry had influenced his early work on quantum mechanics. And Roger Penrose, the eminent mathematician and scientist, was the perfect choice as the seminar's moderator. If anyone could prise the story out of Dirac, it was Penrose.

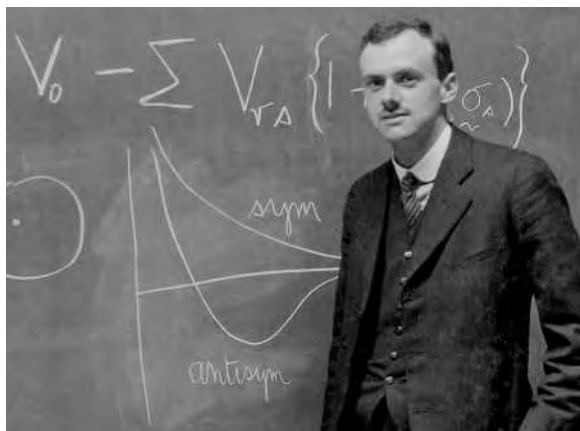
But after giving a short, clear presentation on basic projective geometry, Dirac stopped before connecting it to the quantum world. So once Dirac had answered a few straightforward questions, the disappointed Penrose gently turned to him and asked him point-blank how this geometry talk had influenced his early quantum work. Dirac gave his trademark shake of the head, and declined to speak. Realizing that he was on a hiding to nothing, Penrose had no choice but to fill in the time by extemporizing a short talk on a different subject.

Dirac died 12 years later, still having never clarified the point. Perhaps it was beyond clarification? Perhaps even Dirac did not fully understand the connection between his private geometry and his public algebra? As Ludwig Wittgenstein insisted at the end of his *Tractatus Logico-Philosophicus*, “Whereof one cannot speak, thereof one must remain silent.” ■

Graham Farmelo is Senior Research Fellow at the Science Museum, and Adjunct Professor of Physics at Northeastern University, Boston, USA.

FURTHER READING

Galison, P. *The Suppressed Drawing: Paul Dirac's Hidden Geometry* 145–166 (Univ. California Press, 2000).
Darrigol, O. in *From c-numbers to q-numbers* Ch. 6 (Univ. California Press, 1992).



Good with numbers: Dirac was famed for his skills as an algebraist.

was originally discovered by the Arabs. Dirac remarked that his own preference was “strongly on the side of geometry, and has always remained so”.

Dirac often said that when he was developing quantum mechanics he used his favourite branch of mathematics — projective geometry — which concerns the relationships between points and straight lines. But why then did he not mention geometry in his early papers?

He expunged his geometrical thinking from his early work, he often said in the 1970s, because it was especially expensive in those days for journals to print diagrams and because he thought most physicists were unfamiliar with projective geometry. Only later, when John von Neumann invented ‘state vectors’ did the geometric content of quantum mechanics become plain.

So what were the geometric origins of Dirac's quantum mechanics? I believe that the distinguished French historian of science Olivier Darrigol was the first to set out in detail a plausible explanation. He pointed out that Dirac had his quantum epiphany when he came across an embarrassed

AMERICAN INSTITUTE OF PHYSICS/SPL

NEWS & VIEWS

NANOTECHNOLOGY

Nano-oscillators get it together

Pritiraj Mohanty

Synchronized radiation from arrays of oscillators is widely used in microwave and wireless communications. Phase-locked oscillations produced at the atomic level now pave the way for devices on the nanoscale.

Seen in southeast Asia, it is one of the most dazzling natural visual effects known: large congregations of fireflies blinking on and off in unison (Fig. 1). They orchestrate their flashing in almost perfect rhythm, and at a constant tempo. Each firefly maintains its steady beat through an internal clock, essentially a tiny oscillator inside its brain. Following outside stimuli, this oscillator begins to lock phase, or synchronize, with the firefly congregation¹. A similar thing happens in the human heart: there, a cluster of pacemaker cells, known as the sinoatrial node, generates a synchronous oscillation that commands the rest of the heart to beat, in rhythm, for the duration of a life — typically some three billion pulses. Writing in this issue, Kaka and colleagues (page 389)² and Mancoff and colleagues (page 393)³

report the first demonstration of synchronized oscillation on the nanoscale: the phase-locking of two nano-oscillators in close proximity, through what is known as the spin-torque effect.

Spin is an intrinsic property of a particle or atom, and it is associated with angular momentum. A change in spin state therefore generates a change in angular momentum, resulting in a torque⁴. Use of this phenomenon to find the angular momentum of the photon was proposed by Albert Einstein and Wander de Haas⁵ in 1915, and was achieved experimentally by Richard Beth 20 years later⁶. Since then, various fundamental measurements — notably those of the ratio of angular momentum to magnetic moment (the gyromagnetic ratio) of a metal⁷, and the quantum of superconducting flux⁸ — have relied on spin-torque effects. New approaches to spin-based electronics using mechanical nano-oscillators have been proposed on the strength of the idea⁹. And spin-torque effects have also been discovered^{10,11} in nanoscale magnetic multilayers, allowing steady microwave power to be generated in response to a direct current.

The operational principle of the spin-transfer device used by Kaka *et al.*² and Mancoff

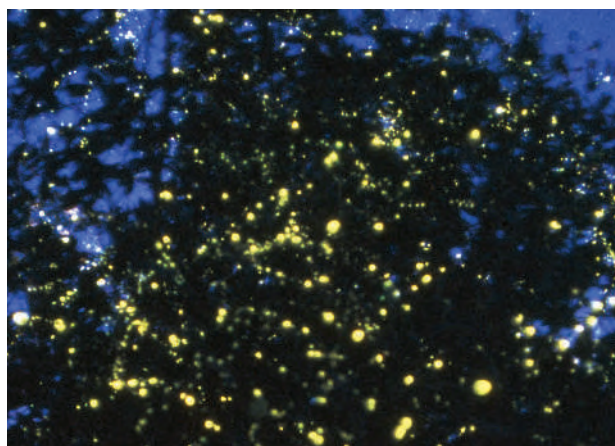


Figure 1 | Fireflies, fireflies burning bright. In the forests of the night, certain species of firefly flash in perfect synchrony — here *Pteroptyx malacciae* in a mangrove apple tree in Malaysia. Kaka *et al.*² and Mancoff *et al.*³ show that the same principle can be applied to oscillators at the nanoscale.

*et al.*³ is well known. This device consists of an electrical point contact linked to multiple thin layers of magnetic material. When a direct current is applied to this contact, torque from the spins of the electrons in the material causes the direction of magnetization to oscillate at microwave frequencies. A spin-transfer oscillator would be expected to produce 'spin-waves', emanating from the region beneath the point contact as each layer of the material influences the next. A second point contact, or a spin-transfer device in close proximity, should experience this spin-wave, leading to phase-locking of the two oscillators — in much the same way that two pendulum clocks coupled through a wall will lock phase, a fact first noted by Christiaan Huygens in the seventeenth century¹².

And here lies the exciting aspect of the latest experiments^{2,3}. Mancoff and co-workers³ vary the distance between the contacts of two identical spin-transfer oscillators and find that, when it is less than roughly 200 nanometres, the oscillators synchronize at a single resonance-peak frequency. Oscillators with a larger inter-contact spacing (typically 400 nanometres) produce two separate resonance peaks, one for each oscillator. The power

radiated from the two locked oscillators is twice that produced from two oscillators at a greater separation radiating independently. Such an enhancement of output power, proportional to the square of the number of oscillators (N^2), is the tell-tale sign of coherent radiation (in the incoherent case, the dependency is on N).

Kaka and co-workers² also find that the power radiated from their device, which consists of two phase-coupled oscillators with different individual power outputs, is consistent with that expected for two phase-coherent signals interfering constructively. Again, this is almost twice that expected from two oscillators radiating at the same frequency but out of phase. In a further testimony to the phase coherence between the

oscillators, the authors find that, as expected, the spread in frequency of the oscillation is reduced in the phase-locked state.

Such phase-locked nano-oscillators^{2,3} have major implications for the use of nanoscale spin-transfer devices. The output power of a single device is small (typically less than a millionth of a milliwatt), but connecting two or more phase-locked devices together could quickly increase the output to a useful level of the order of microwatts or even milliwatts at gigahertz frequencies. The radiation pattern produced by an array of oscillators vibrating in phase is highly directional, making them useful as beam-steering devices in wireless communications — as either transmitters or receivers. Before such a device can be used on the nanoscale, however, phase-locking among many nano-oscillators must be demonstrated.

Finally, the significance of the oscillators' spatial distribution adds an exciting dimension to the problem. It creates the potential for probing synchronization and chaos at the nanoscale, an active field of research in applied mathematics and neuroscience. Motivation for future work here can once again be found in the stunning visual patterns of the spatial temporal dynamics of fireflies. Nature never fails to inspire. ■

I. POLUNIN/NHPA

fails to inspire.

Pritiraj Mohanty is in the Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA.

e-mail: mohanty@physics.bu.edu

1. Strogatz, S. *Syn: The Emerging Science of Spontaneous Order* (Hyperion, New York, 2003).
2. Kaka, S. *et al. Nature* **437**, 389–392 (2005).
3. Mancoff, F. B. *et al. Nature* **437**, 393–395 (2005).

4. Richardson, O. W. *Phys. Rev.* **26**, 248–253 (1908).
5. Einstein, A. & de Haas, W. J. *Verh. Dt. Phys. Ges.* **17**, 152–170 (1915).
6. Beth, R. A. *Phys. Rev.* **50**, 115–125 (1936).
7. Scott, G. G. *Phys. Rev.* **82**, 542–547 (1951).
8. Doll, R. & Nabauer, M. *Phys. Rev. Lett.* **7**, 51–52 (1961).
9. Mohanty, P. *et al. Phys. Rev. B* **70**, 195301 (2004).
10. Berger, L. *Phys. Rev. B* **54**, 9353–9358 (1996).
11. Slonczewski, J. C. *J. Magn. Magn. Mater.* **159**, L1–L7 (1996).
12. Huygens, C. *Oeuvres Complètes de Christiaan Huygens* (ed. Nijhoff, M.) (Société Hollandaise des Sciences, The Hague, 1893).

GENOMICS

Massively parallel sequencing

Yu-Hui Rogers and J. Craig Venter

A sequencing system has been developed that can read 25 million bases of genetic code — the entire genome of some fungi — within four hours. The technique may provide an alternative approach to DNA sequencing.

Since the publication of the first complete genome sequence of a living organism¹ in 1995, the field of genomics has changed dramatically. Fuelled by innovations in high-throughput DNA sequencing, high-performance computing and bioinformatics, genomic science has expanded substantially and the rate of genomic discovery has grown exponentially. To date, the genomes of more than 300 organisms have been sequenced and analysed, including those of most major human pathogens, diverse microbes — and, of course, our own genome^{2,3}. These advances have profoundly altered the landscapes of biological science and medicine. In this issue, Rothberg and colleagues (page 376)⁴ describe a sequencing system that offers a much higher throughput than the current state-of-the-art methods. The system has some limitations to overcome before it can be used for all sequencing applications, but it is nonetheless one of the most promising sequencing technologies to have emerged in recent years.

For more than a decade, Sanger sequencing⁵ and fluorescence-based electrophoresis technologies⁶ have dominated the DNA sequencing field. Continued improvements in these techniques and in instrumentation, paired with advances in computing and informatics, have reduced the cost of sequencing by roughly two orders of magnitude and transformed genome projects from decade-long endeavours to projects of mere months (for mammalian-sized genomes), or even weeks (for microbial genomes). However, it still costs an estimated US\$10 million to US\$25 million to sequence a single human genome⁷ and \$20,000–\$50,000 to sequence a microbial genome. Only a handful of large genome centres worldwide have the resources and technical expertise to handle the sequencing of a mammalian-sized genome, perform large-scale sequencing of multiple organisms or conduct the resequencing of large numbers of genes. To ensure continued growth of genomic science and to enable more

labs to become involved in DNA sequencing, new approaches must decrease the cost and increase the throughput of sequencing significantly, while maintaining the high quality of data produced by the current approach.

Rothberg and colleagues⁴ have developed a highly parallel system capable of sequencing 25 million bases in a four-hour period — about 100 times faster than the current state-of-the-art Sanger sequencing and capillary-based electrophoresis platform. The method could potentially allow one individual to prepare and sequence an entire genome in a few days (Fig. 1). The sequencer itself, equipped with a simple detection device and liquid delivery system, and housed in a casing roughly the size of a microwave oven, is actually relatively low-tech. The complexity of the system lies primarily in the sample preparation and in the microfabricated, massively parallel platform, which contains 1.6 million picolitre-sized reactors in a 6.4-cm² slide.

Sample preparation starts with fragmentation of the genomic DNA, followed by the attachment of adaptor sequences to the ends of the DNA pieces. The adaptors allow the DNA fragments to bind to tiny beads (around 28 µm in diameter). This is done under conditions that allow only one piece of DNA to bind to each bead. The beads are encased in droplets of oil that contain all of the reactants needed to amplify the DNA using a standard tool called the polymerase chain reaction. The oil droplets form part of an emulsion so that each bead is kept apart from its neighbour, ensuring the amplification is uncontaminated. Each bead ends up with roughly 10 million copies of its initial DNA fragment.

To perform the sequencing reaction, the DNA-template-carrying beads are loaded into the picolitre reactor wells — each well having space for just one bead. The technique uses a sequencing-by-synthesis⁸ method developed by Uhlen and colleagues, in which DNA

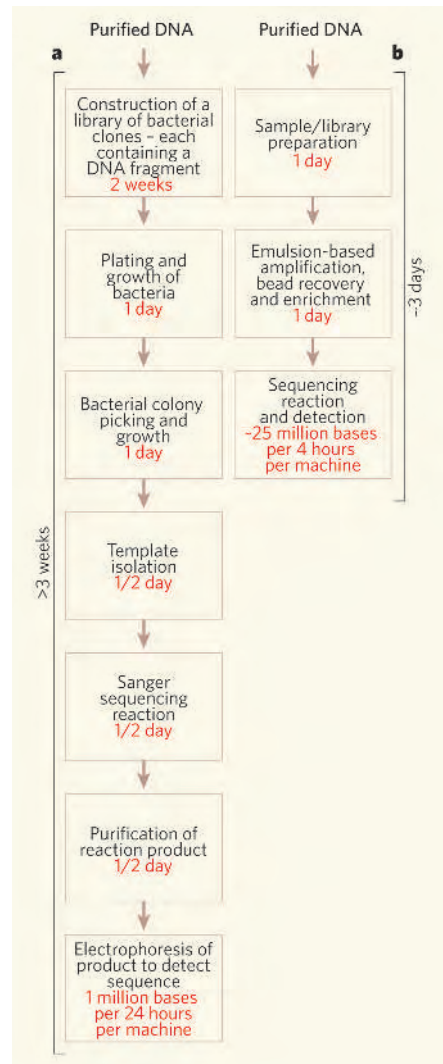


Figure 1 | Speeding up sequencing. Flow diagrams for **a**, traditional microlitre-scale Sanger DNA sequencing and electrophoresis, and **b**, the massively parallel picolitre-scale sequencing developed by Rothberg *et al.*⁴. The traditional microlitre-scale approach requires a longer processing time per production cycle, substantially more support equipment, a larger facility and more labour than the picolitre-scale approach.

complementary to each template strand is synthesized. The nucleotide bases used for sequencing release a chemical group as the base forms a bond with the growing DNA chain, and this group drives a light-emitting reaction in the presence of specific enzymes and luciferin. Sequential washes of each of the four possible nucleotides are run over the plate, and a detector senses which of the wells emit light with each wash to determine the sequence of the growing strand.

This new system shows great promise in several sequencing applications, including re-sequencing and *de novo* sequencing of smaller bacterial and viral genomes. It could potentially allow research groups with limited resources to enter the field of large-scale DNA sequencing and genomic research, as it provides a technology that is inexpensive and easy to implement and maintain. However, this technology

cannot yet replace the Sanger sequencing approach for some of the more demanding applications, such as sequencing a mammalian genome, as it has several limitations.

First, the technique can only read comparatively short lengths of DNA, averaging 80–120 bases per read, which is approximately a tenth of the read-lengths possible using Sanger sequencing. This means not only that more reads must be done to cover the same sequence, but also that stitching the results together into longer genomic sequences is a lot more complicated. This is particularly true when dealing with genomes containing long repetitive sequences.

Second, the accuracy of each individual read is not as good as with Sanger sequencing — particularly in genomic regions in which single bases are constantly repeated. Third, because the DNA 'library' is currently prepared in a single-stranded format, unlike the double-stranded inserts of DNA libraries used for Sanger sequencing, the technique cannot generate paired-end reads for each DNA fragment. The paired-end information is crucial for assembling and orientating the individual sequence reads into a complete genomic map for *de novo* sequencing applications. Finally, the sample preparation and amplification processes are still quite complex and will require automation and/or simplification.

Church and colleagues⁹ also recently hit upon the idea of using massively parallel reactions to speed up sequencing, although their method is still only at the proof-of-principle stage rather than being a full production system. They use a similar principle to Rothberg and colleagues⁴, that is, sequencing-by-synthesis on a solid support. However, the two approaches diverge in terms of library construction, sequencing chemistry, signal detection and array platform. These differences greatly affect the characteristics and reproducibility of the data, as well as the scalability of the platform. For example, Church and colleagues' method can read paired-end sequences; however, its average read-lengths are approximately a fifth of those generated by Rothberg and colleagues' system. These differences are key factors in determining the sequencing application for which each technique might be most suited.

It may be years before Rothberg and colleagues' system, or other similar approaches^{9,10}, can tackle all three billion letters of the human genome with the same reliability and accuracy as current methods. Nevertheless, it looks extremely promising, and it is certainly one of the most significant sequencing technologies under development.

Yu-Hui Rogers and J. Craig Venter are at the J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, Maryland 20850, USA.
e-mail: jcventer@venterinstitute.org

- Margulies, M. *et al.* *Nature* **437**, 376–380 (2005).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. Natl Acad. Sci. USA* **74**, 5463–5467 (1977).
- Prober, J. M. *et al.* *Science* **238**, 336–341 (1987).
- NIH News Release www.genome.gov/12513210 (2004).

- Nyren, P., Pettersson, B. & Uhlen, M. *Anal. Biochem.* **208**, 171–175 (1993).
- Shendure, J. *et al.* *Science* advance online publication doi:10.1126/science.1117389 (2005).
- Quake, S. R. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 3960–3964 (2003).

DEVICE PHYSICS

Enlightening solutions

Klaus Meerholz

White-light-emitting diodes are becoming increasingly important, but what is the best way to build compact devices possessing high efficiency? Bright prospects are offered by multi-layer organic devices grown from solution.

Sources of white light are found almost everywhere — in lighting and signage generally, but increasingly as backlights in all sorts of displays, for example in laptop computers or smart phones. The development of organic light-emitting diodes (OLEDs) promises further innovation in the field: such diodes are lightweight, provide high brightness at low power, and can be fabricated on flexible substrates to form thin devices at potentially low production cost¹. Writing in *Advanced Materials*, Gong *et al.*² demonstrate the use of semiconducting electrolytes to grow highly efficient, multi-layer, white-light OLEDs from solution.

The structure of an OLED is quite simple; it uses an organic material that either fluoresces or phosphoresces. (Both processes involve the re-emission of absorbed light at a longer wavelength; in phosphorescence, the quantum-mechanical processes that lead to re-emission are more complex, so the emission is delayed.) The light-emitting material is sandwiched as a thin film, typically 70–100 nanometres thick, between two electrodes. Of these, the anode is typically transparent and the cathode acts as a mirror that ideally reflects any incident photons back towards the transparent side.

When a voltage is applied to the electrodes, positive and negative charges ('holes' and electrons, respectively) are injected into the film and move towards each other, forming a body known as an exciton on meeting. This exciton can become de-excited by emitting a photon, which leaves the device through the transparent anode. Excitons are divided into two categories according to the alignment of the spins of the electrons involved relative to one another: if these are antiparallel, a 'singlet' with a total spin of zero is formed; if they are parallel, the state is a 'triplet' with a spin of one. In terms of quantum mechanics, three-quarters of excitons must be triplets and only one-quarter singlets. The prospect of achieving higher efficiency with OLEDs emitting from triplets has led to the investment of considerable effort^{2,3} in their development.

OLEDs are generally produced by one of two routes: the sublimation of small molecules

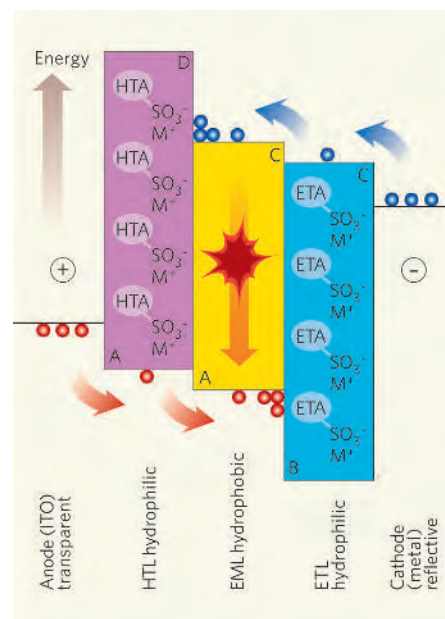


Figure 1 | Cross-section through the multi-layer OLED developed by Gong and colleagues².

The staggered height of the layers indicates their different energies; electrons (blue) favour moving to lower energies, whereas holes (red) tend towards higher energies. The emissive layer (EML) is sandwiched between a hole-transport layer (HTL) and an electron-transport layer (ETL) consisting of transport agents (HTA/ETA) containing sulphonate groups (SO_3^-). (M^+ stands for a metal counter-ion.) This structure serves three purposes: first, to facilitate the injection of holes and electrons (A and C) into the emissive layer by reducing the energetic barriers to their passage; second, to enhance the recombination efficiency (formation of excitons) by blocking the passage of one type of carrier (B or D) from the emissive layer into the opposite transport layer through a large step in energy; and third, to avoid quenching reactions of excitons at the electrodes (+/−). The layers of the OLED are deposited alternately from water or ethanol (HTL and ETL) and from organic solvents (EML). The light emitted through recombination in the EML passes through the transparent anode to the left; the colour of the emission depends on the material or mixture of materials that forms the EML.

- Venter, J. C. *et al.* *Science* **269**, 496–512 (1995).
- Venter, J. C. *et al.* *Science* **291**, 1304–1351 (2001).
- International Human Genome Mapping Consortium. *Nature* **409**, 860–921 (2001).



50 YEARS AGO

The Spirit of Liberty: Papers and Addresses of Learned Hand — The papers and addresses collected in this book range for the most part over fields with which the man of science is concerned rather as citizen than as scientist...Supremely the book testifies to the truth of Judge Hand's contention that there is no substitute for an open mind enriched by reading and the arts, and that the scientific worker who would influence the thought and action of his time must have some acquaintance, the wider the better, with what others have thought and felt in circumstances as near as possible to those of the groups in question...The plea for the open mind and free discussion in which these words occur is much more than championship of dissent and a direct rebuke to the excesses of McCarthyism and secrecy and security procedures: it is an outstanding statement of the case for general education as the basis of a free society.

From *Nature* 17 September 1955.

100 YEARS AGO

At the present time by far the most serious problem which the automobilist has to face is the abatement of the "dust nuisance". A great deal of bad feeling has arisen against the motorist on account of the dust which he too frequently produces, and there is no doubt that there are very good grounds for the irritation which has arisen, particularly in agricultural districts...Although a permanently good road may be made by the use of materials such as Tarmac, and dusty roads may be cured temporarily by various means, yet such measures can be taken only over a small proportion of our roads owing to the cost. In towns and large villages the roads might be suitably treated, but the average motorist seeks the country, and the greater part of the routes which he wishes to traverse will not pay for any special treatment.

From *Nature* 14 September 1905.

in a vacuum, or the wet-chemical deposition of polymers onto a substrate. Which of these approaches will ultimately survive in a production environment is still a matter of debate. Although the small-molecule approach improved device performance, especially device lifetime, wet-chemical deposition is the more attractive technique for mass production. This is because it allows layers of polymers to be laid down cheaply in a roll-to-roll process using common techniques such as those used for screen and inkjet printing. It is therefore the method of choice for Gong *et al.*² and many others.

Multi-layer OLEDs are generally more efficient than single-layer types. In the most efficient OLEDs, the emission layer is sandwiched between a hole-transport layer and an electron-transport layer (Fig. 1). Fabricating such multi-layer structures from solution is challenging (vacuum deposition of small molecules is relatively straightforward). It is crucial to ensure that layers already deposited from solution are totally resistant to the solvents used to deposit subsequent layers, to avoid intermixing.

There are three main ways to do this. The first is to use 'orthogonal' solvents for the individual layers — that is, the solvent used in one deposition does not dissolve any previous layer. For example, the conductive polymer poly(3,4-ethylenedioxy)thiophene (PEDOT), commonly used for OLED anodes, is deposited from an aqueous suspension. After drying, further organic layers can be deposited from typical organic solvents such as toluene without redissolving the PEDOT. A second method is to change the polarity or solubility of the deposited material. An example here is the first luminescent polymer ever discovered, poly(*p*-phenylenevinylene), or PPV (ref. 4), where a polar sulphonium precursor molecule is transformed by heating into a nonpolar polymer that is insoluble in all organic solvents.

A third, highly attractive approach is to introduce several reactive molecular groups into a semiconductor material. These can be polymerized after deposition to yield totally insoluble crosslinked layers, a process that can, in principle, be repeated indefinitely. In recent years, many materials that can be processed from solution and possess the ability to form multi-layers have been proposed, the most promising being oxetanes⁵, styrenes⁶, dienes⁷ and trifluorovinyl ethers⁸. The efficiency of OLEDs at incorporating such materials (most of which are hole-transporters) is in many cases greater than that for reference devices using just PEDOT as anode, or the transparent metal indium tin oxide (ITO).

Gong *et al.*² propose an extension of the first, orthogonal-solvent approach. They developed derivatives of two commonly used organic semiconducting materials — the hole-conducting poly(*N*-vinylcarbazole), or PVK, well known from the early days of xerography, and the electron-conducting oxadiazole derivative

PBD. The authors achieved this by incorporating into them ionic sulphonate groups, which make the derivatized material soluble in highly polar solvents such as water and ethanol but insoluble in organic solvents. This trick allowed them to create the layers using aqueous or ethanol solution, as the components of the emissive layer of the OLED were totally insoluble in either solvent. (The emissive layer contained a fluorescent polymer emitting green and blue light, doped with a phosphorescent molecule that sends out red light, yielding an overall white emission.)

By alternating deposition from hydrophilic and hydrophobic solvents, Gong and colleagues built a three-layer device (Fig. 1). Because the emissive layer acts as a barrier against the redissolution of the first deposited layer (the hole-transport layer) during the deposition of the third layer (the electron-transport layer), it was of crucial importance that all layers were free of pinholes. The resulting OLEDs produced a luminous intensity of around 10 candelas per ampere of supplied current — around 2.5 lumens per watt. The efficiency of the device is thus among the highest to date for solution-processed white-light-emitting devices¹, and one-and-a-half to three times better than reference devices in which either of the two transport layers was missing.

It might be thought that the introduction of sulphonate groups and the consequent presence of mobile metal counter-ions might impair the device's performance. Certainly, devices using transport layers show slightly increased onset voltages — the minimum supply voltage at which emission will occur. According to Gong and colleagues, this is due to the greater thickness of the device compared with other OLEDs and could thus presumably be improved by adjusting the layer thicknesses, or by using different hole- and electron-transport materials containing sulphonate ions or other ionic groups.

The big question, however, is whether devices based on such sulphonate materials can reach the operating lifetimes necessary for practical applications (typically more than 10,000 hours). The presence of mobile metal ions could cause similar problems here to those seen with electrochemical emissive devices. Although such questions remain unresolved, Gong and colleagues' contribution² is a step towards a more flexible, lower-cost source of white light. ■

Klaus Meerholz is at the Institute of Physical Chemistry, University of Cologne, Luxemburgerstrasse 116, 50939 Cologne, Germany. e-mail: klaus.meerholz@uni-koeln.de

1. Andrade, B. W. & Forrest, S. R. *Adv. Mater.* **16**, 1585–1595 (2004).
2. Gong, X. *et al.* *Adv. Mater.* **17**, 2053–2058 (2005).
3. Baldo, M. A. *et al.* *Nature* **395**, 151–154 (1998).
4. Burroughes, J. H. *et al.* *Nature* **347**, 539–541 (1990).
5. Müller, C. D. *et al.* *Nature* **421**, 829–833 (2003).
6. Klärner, G. *et al.* *Chem. Mater.* **11**, 1800–1807 (1999).
7. Aldred, M. P. *et al.* *Chem. Mater.* **16**, 4928–4936 (2004).
8. Gong, X. *et al.* *Appl. Phys. Lett.* **83**, 183–185 (2003).

EARTHQUAKES

Giant returns in time

Sergio Barrientos

The behaviour of a seismic fault in Chile seemed to confound predictions of how often giant earthquakes should recur. Examination of a 2,000-year record of tsunami deposits in the region clarifies matters.

In May 1960, south-central Chile experienced a huge earthquake, the largest since instrument records began. The consequences were felt not only in Chile, but also in Hawaii, the Philippines, Japan and other locations across the Pacific, which were all hit by the ensuing tsunami. But this giant seismic event has long puzzled seismologists, because the energy released by the earthquake should have taken several centuries to build up, through the accumulation of stress on the fault concerned, as the Nazca and South American tectonic plates converge at a rate of 8–9 centimetres per year. Yet the immediately preceding events are historically documented as occurring not only in 1575, which would fit expectations, but also in 1737 and 1837.

As they describe on page 404 of this issue¹, Cisternas *et al.* have revisited this puzzle by examining records of land movement and tsunamis left in buried layers of sand and soil in an estuary that runs across the central part of the fault zone (see Fig. 1 of the paper, page 404). With these records, and earlier ones (Fig. 1), the history of activity on this fault can be rewritten.

Since the inception of instrumental recording in the late 1800s, the largest earthquakes have all occurred in subduction zones², where one tectonic plate is being driven beneath another. These earthquakes involve ruptures on the order of a thousand kilometres long by a couple of hundred kilometres wide, with fault displacements of tens of metres. Hundreds of years are necessary to accumulate the stresses released in these giant events, which leave traces of their consequences that can later be identified.

Thanks to the work of Cisternas *et al.*¹, we now have more insight into the seismic cycle in south-central Chile. The earthquake of May 1960 resulted from a rupture, about 1,000 km long and 150 km wide, along the north–south-trending fault where the Nazca plate dives beneath South America. Earthquakes of this size saturate the recorders of the seismic waves that are usually used to estimate earthquake magnitude and which are reported in terms of the familiar Richter scale. Instead, a measure known as seismic moment is used³. This is now more commonly applied to seismic events in general, and is calculated by multiplying the area of the rupture zone by the fault displacement and a quantity reflecting the rigidity of the volume in which the rupture takes place. The moment magnitude, derived from this quantity, reflects the real size in terms of the elastic energy release of an earthquake.

The 1960 earthquake measured 9.5 on the moment-magnitude scale. It has a special place in seismological history, because it provided experimental confirmation of the idea that earthquakes can cause free oscillations of the Earth⁴ — that is, set the Earth ringing. The observed changes in land elevation ranged from 6 m of uplift to 2 m of subsidence; these displacements have been modelled^{5,6} as the elastic response of the Earth to an average dislocation of 20 m along the fault, with localized peaks of more than 30 m. Even now, in 2005, post-seismic readjustments continue to be observed in the area⁷.

Cisternas *et al.*¹ undertook a detailed examination of a local stratigraphic record of earthquakes in south-central Chile, as produced by

the associated tsunamis and seen in layers of intercalated sand and soil extending horizontally for nearly a kilometre close to the mouth of the Río Maullín. This region subsided approximately 2 m as a consequence of the 1960 event, but the stratigraphy has been preserved because of long-term net tectonic uplift. Overall, the authors were able to identify a sequence of eight large earthquakes that have occurred over the past 2,000 years, with an average recurrence interval of about 300 years.

But what about the apparent recurrence sequence of 1575, 1737, 1837 and 1960, which does not fit this pattern? The consequences of the 1575 event are evident both from carbon dating and in documents written by the Spanish conquistadors, which tell of effects that were similar to those seen in 1960. Documentation of the earthquakes of 1737 and 1837 is less clear, but they are frequently cited in the seismological histories of the region as being large (estimated magnitudes of 7.5 and 8.0, respectively)⁸.

Cisternas *et al.*, however, find that although the 1575 earthquake appears clearly in their tsunami stratigraphic record, those of 1737 and 1837 do not. A tsunami was associated with the event of 1837, reaching Hawaii with an amplitude of 6 m. But Cisternas *et al.* suggest that it originated outside their study area, to the south, and they conclude that the earthquake of 1737 was too small to generate a sizeable tsunami. All in all, they propose that much of the fault dislocation that occurred in 1960 stemmed from a release of energy that had remained 'locked in' since 1575 — that is, the earthquakes that occurred between those times had expended comparatively little of the accumulated stress.

The relevance of these studies is underlined by the more recent occurrence of a huge earthquake elsewhere in the world. In December 2004, and again in March 2005, the eastern Indian Ocean was seriously affected by tectonic movement at a subduction zone. The giant December event stemmed from a displacement of up to 30 m along a rupture 1,100 km in length⁹, which produced the devastating tsunami that swept across the Indian Ocean — a startling reminder of the need to learn more about the behaviour of giant earthquakes. ■

Sergio Barrientos is at the Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization, PO Box 1200, A-1400 Vienna, Austria. e-mail: sergio.barrientos@ctbto.org



Figure 1 | Spot the surveyor. In this print, produced in 1874, the figure of Francisco Vidal Gormaz can be seen at work in the coastal village of Carelmapu in southern Chile. Here he surveys part of the region that will be hit by the 1960 earthquake; his records have helped in interpreting the earthquake.

1. Cisternas, M. *et al.* *Nature* **437**, 404–407 (2005).
2. Kanamori, H. *J. Geophys. Res.* **82**, 2981–2987 (1977).
3. Aki, K. *Bull. Earthq. Res. Inst. Tokyo Univ.* **44**, 23–88 (1966).
4. Benioff, H., Press, F. & Smith, S. *J. Geophys. Res.* **66**, 605–619 (1961).
5. Plafker, G. & Savage, J. C. *Geol. Soc. Am. Bull.* **81**, 1001–1030 (1970).
6. Barrientos, S. & Ward, S. N. *Geophys. J. Int.* **103**, 589–598 (1990).
7. Klotz, J. *et al.* *Earth Planet. Sci. Lett.* **193**, 437–446 (2001).
8. Lomnitz, C. *Seism. Res. Lett.* **75**, 368–378 (2004).
9. Vigny, C. *et al.* *Nature* **436**, 201–206 (2005).

MICROBIOLOGY

Bacterial speech bubbles

Stephen C. Winans

Many bacteria socialize using diffusible signals. But some of these messages are poorly soluble, so how do they move between bacteria? It seems they can be wrapped up in membrane packages instead.

On page 422 of this issue, Mashburn and Whiteley¹ describe the unexpected convergence of two seemingly unrelated areas of microbiological research: how bacteria talk to their friends, and how they attack their enemies. The authors studied the bacterial pathogen *Pseudomonas aeruginosa*, which releases a hydrophobic molecule called the 'pseudomonas quinolone signal' (PQS) to send messages to other bacteria of the same species. The surprise is that, rather than being secreted as single molecules, PQS is released in bubble-like 'vesicles' that also contain antibacterial agents and probably toxins aimed at host tissue cells as well.

Various groups of bacteria use diffusible chemicals to signal to their own kind, and this method of communication seems to have evolved independently several times². *Pseudomonas aeruginosa* is a Gram-negative proteobacterium and, like many proteobacteria, it secretes molecules called acylhomoserine lactones (AHLs). The concentration of these signal molecules increases with cell population density, enabling the bacteria to sense how many of their kind surround them — a phenomenon known as quorum sensing. Whole sets of genes are switched on only at high cell densities, including genes for bioluminescence, DNA transfer, pigment production, and genes required for infecting plants, animals and humans². But AHL molecules are readily broken down by other bacteria, and some AHL signals are poorly soluble in water, so they cannot travel far in an aqueous environment — both factors that would seem to limit their potential as communication signals. PQS, although not an AHL-type signal, is also poorly water-soluble and readily broken down by other bacteria. So how do these molecules manage to spread the word from one bacterium to another?

Mashburn and Whiteley found clues from another bacterial communication system — one aimed at enemies rather than friends. Many Gram-negative bacteria release minute vesicles (about 50 nm in diameter)³ from their outer membranes as a means of delivering toxins to host cells and other bacteria (sometimes referred to type VI protein export⁴). The vesicles consist of a lipid bilayer surrounding an aqueous core and they can therefore transport lipid-soluble toxins (lipopolysaccharide endotoxin) on their surface and protein toxins in their core. They release their poisonous cargo

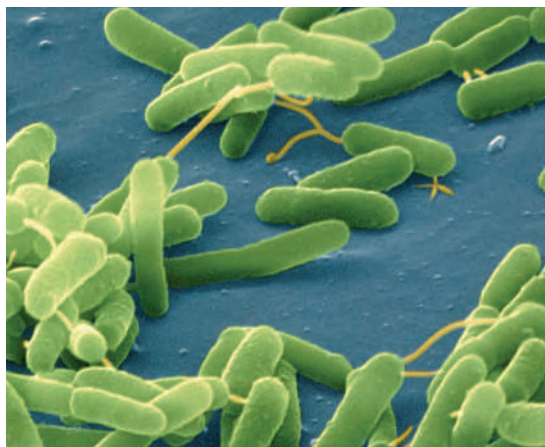


Figure 1 | Friend or foe? Gram-negative proteobacteria such as *Pseudomonas aeruginosa* communicate with each other by secreting chemical messengers. Mashburn and Whiteley¹ show that they also release such signals in vesicles that contain toxins against other bacteria and the cells of their host.

by fusing with the lipid bilayer of target cells. Mashburn and Whiteley show that these vesicles can also release the PQS signal to other bacteria, thereby sidestepping the problems of poor water-solubility and degradation.

Remarkably, *P. aeruginosa* has three quorum-sensing systems that use AHL signals. Apart from the PQS system, there is the LasR/LasI system, which signals using 3-oxo-dodecanoylhomoserine lactone (OdDHL), and the RhlR/RhlI system, which uses butyrylhomoserine lactone (BHL). Of these three signals, OdDHL and PQS are poorly water-soluble, and all three can be degraded by other bacteria. Mashburn and Whiteley found that PQS (but not OdDHL or BHL) is not dispersed between bacteria as single, water-soluble molecules, but rather is concentrated within membrane vesicles, presumably buried within the lipid bilayer.

PQS is one of at least 55 quinolones and quinolines made by *P. aeruginosa*, some of which have potent antibacterial activity. The membrane vesicles, in addition to signalling to kindred *P. aeruginosa* cells, showed potent antibacterial activity against the Gram-positive bacterium *Staphylococcus aureus*. Most of the antimicrobial activity could be extracted by organic solvents, indicating that toxicity was due to quinolones rather than to the protein component of the vesicles.

The production of the vesicles seems to be regulated by PQS, because when PQS synthesis was abolished by a mutation in the *pqsH* gene,

vesicle production virtually ceased. Addition of extra PQS restored vesicle formation, even if protein synthesis was blocked by an antibiotic. So, whatever PQS is doing, it is probably happening at the post-translational level, rather than causing alterations in gene expression. Most of the added PQS was found in vesicles, which also contained other quinolones possessing antimicrobial activity. It seems that when PQS is inserted into the outer membrane — whether it was made by that bacterium or added from another source — it can destabilize the membrane and spontaneously form vesicles containing itself, other quinolones and presumably endotoxin and periplasmic proteins. These vesicles can then signal to other pseudomonads, kill unrelated bacteria, and deliver toxic lipopolysaccharides and proteins to host cells.

Pseudomonas aeruginosa is an opportunistic human pathogen that is a particular problem in the lungs of cystic fibrosis patients. Greater understanding about how these bacteria communicate with each other and attack host cells could lead to novel treatment strategies for *P. aeruginosa* infections. The way in which other bacteria generate these ubiquitous membrane vesicles, and what they might contain, is unknown, but most proteobacteria do not synthesize PQS, so other mechanisms must be involved. Moreover, hydrophobic long-chain AHLs were not found in membrane vesicles in this study, so how these signalling molecules diffuse or are transported between bacteria has still to be discovered.

Stephen C. Winans is in the Department of Microbiology, Cornell University, Ithaca, New York 14853, USA.
e-mail: scw2@cornell.edu

1. Mashburn, L. M. & Whiteley, M. *Nature* **437**, 422–425 (2005).
2. Waters, C. M. & Bassler, B. L. *Annu. Rev. Cell Dev. Biol.* (in the press)
3. Kadurugamuwa, J. L. & Beveridge, T. J. *J. Antimicrob. Chemother.* **40**, 615–621 (1997).
4. Wai, S. N. *et al.* *Cell* **115**, 25–35 (2003).

Correction

In the News & Views article "Cell biology: Powerful curves" by L. Mahadevan and T. J. Mitchison¹, which discussed a paper published in the same issue², the authors put forward an elastic-sheet model of microtubule structure. Such a model was previously proposed³ and applied^{3–5} by Flyvbjerg and colleagues in earlier papers. The authors apologize for the oversight in not citing these earlier quantitative studies.

1. Mahadevan, L. & Mitchison, T. J. *Nature* **435**, 895–897 (2005).
2. Wang, H.-W. & Nogales, E. *Nature* **435**, 911–915 (2005).
3. János, I. M., Chrétien, D. & Flyvbjerg, H. *Eur. Biophys. J.* **27**, 501–513 (1998).
4. Chrétien, D., János, I. M., Taveau, I. & Flyvbjerg, H. *Cell Struct. Funct.* **24**, 299–303 (1999).
5. János, I. M., Chrétien, D. & Flyvbjerg, H. *Biophys. J.* **83**, 1317–1330 (2002).

OBITUARY

Charles David Keeling 1928–2005

Pioneer in the modern science of climate change.

Numerous records now show how we humans are altering the planet, with potentially global consequences for climate. But the first and now iconic examples of documenting global climate change were the precise measurements of the concentration of carbon dioxide in the atmosphere made at the mountain station on Mauna Loa, Hawaii, by Charles David Keeling. If the world today realizes that it has a problem and needs to curb emissions of greenhouse gases, it in large part owes this knowledge to Keeling's painstaking efforts.

Keeling was born in Pennsylvania in 1928, and studied chemistry and isotope geochemistry in Chicago. After completing his PhD, in 1953 he moved to the California Institute of Technology, and became interested in the problem of measuring levels of CO₂ in the atmosphere. Before that time, the prospect that levels of atmospheric CO₂ could be altered on a global scale because of the burning of coal, gas and oil was largely theoretical. At the end of the nineteenth century, Svante Arrhenius and Arvid Högbom had suggested that temperature might be affected by greenhouse gases. But the possibility of increasing atmospheric CO₂ was of no concern because the ocean was believed to absorb most of any increase.

In the 1950s, however, Roger Revelle and Hans Suess at the Scripps Institution of Oceanography in California realized that the sink capacity of the ocean is limited because of the slow mixing with deeper waters. Measurement of atmospheric CO₂ thus became a compelling topic of scientific interest. A job offer from Revelle brought Keeling to Scripps, where he remained for his entire career.

Keeling looked for alternatives to the existing, questionable methods of determining variations in CO₂, and eventually resorted to using an infrared analyser. As this is a relative measurement, he had to devise a clever system of calibration to make the measurements accurate and comparable between different sites and over long time spans. In conjunction with the International Geophysical Year 1957, a series of global geophysical activities spanning 1957–58, systems were set up on Mauna Loa and in Antarctica, supplemented by lab measurements on air flasks regularly sampled at the South Pole. The choice of such remote places was to ensure that they were as far away from local sources of CO₂ as possible, thus allowing reliable detection of changes in the background atmosphere.

Only two years later, Keeling produced a classic paper showing that the atmospheric CO₂ concentration was rising at the South Pole at a rate consistent with estimates of fossil-fuel emissions and ocean uptake. He also showed that the concentration at Mauna Loa exhibits a characteristic seasonal variation. Using concurrent measurements of the ¹³C/¹²C isotope ratio in CO₂, which differs according to the exchange processes with the land and the ocean, Keeling was able to demonstrate that this seasonal variation is probably driven by the annual cycle of vegetation in the northern extratropics.

Many would have left it at that and moved on to other scientific investigations. Keeling, however, anticipated the power of long-term observations and, despite many funding crises, relentlessly pursued the measurements. Thanks to this work, we now have two highly accurate records of atmospheric CO₂ spanning almost half a century; between the late 1950s and today, levels of CO₂ have risen from about 315 parts per million to more than 375 parts per million.

The information content in these long records is truly amazing. Keeling and his co-workers showed that the changes in total atmospheric CO₂ content can be estimated quite accurately. And by combining this with estimates of fossil-fuel emissions, we can tell how much CO₂ is being absorbed jointly by the ocean and terrestrial biosphere. Close analysis of the seasonal cycle in CO₂ concentrations at Mauna Loa revealed an increase in amplitude of up to 20% over the 40 years, indicating a strengthening 'breathing' of the vegetation in the Northern Hemisphere. The records also show interannual variations, which can be related to climate fluctuations such as those of the El Niño–Southern Oscillation. And using concurrent ¹³C/¹²C isotope measurements, Keeling's group could estimate the relative contributions of land and ocean to these phenomena.

The records from Mauna Loa and the South Pole were soon complemented by measurements made elsewhere to produce a global picture of CO₂ distribution. Keeling was aware that these data could be used with a model of atmospheric transport to determine the regional patterns of CO₂ sources and sinks. During a sabbatical stay in Stockholm in the early 1960s, he and the meteorologist Bert Bolin performed the first such 'inversion' of atmospheric CO₂



measurements and demonstrated that fossil-fuel emissions indeed contribute significantly to the global distribution of the gas. As increasing numbers of atmospheric monitoring stations became established by other groups, this pioneering work was refined with better numerical models. Today it constitutes a key method for monitoring the carbon cycle on larger scales.

Keeling was not always an easy colleague to work with. He was meticulous and demanded the same standards from his co-workers. But it is thanks to this attention to detail that we now have such beautiful documents depicting the changes in atmospheric composition. His work was recognized by numerous awards, such as the Blue Planet Prize from the Science Council of Japan and the Tyler Prize for Environmental Achievement.

Outside science, Keeling led an active life. Supported by his lovely family, he was involved in the political life of his home town; he also conducted a choir, and after a hard day at work he enjoyed playing the grand piano (and did so with considerable proficiency). He took great interest in foreign cultures — I fondly remember how, during a sabbatical stay in Switzerland, he even took on the challenge of taking lessons in the Bernese Swiss–German language with considerable success. But his major love was nature, and especially the mountains. Keeling died on 20 June. Appropriately, his final resting place is near his summer house in the wilderness of Montana. ■

Martin Heimann

Martin Heimann is at the Max Planck Institute for Biogeochemistry, PF 100164, 07701 Jena, Germany.
e-mail: martin.heimann@bgc-jena.mpg.de

SCRIPPS INST. OCEANOGR./UCSD

BRIEF COMMUNICATIONS

Age written in teeth by nuclear tests

A legacy from above-ground testing provides a precise indicator of the year in which a person was born.

Establishing the age at death of individuals is an important step in their identification and can be done with high precision up to adolescence by analysis of dentition, but it is more difficult in adults. Here we show that the amount of radiocarbon present in tooth enamel as a result of nuclear bomb testing during 1955–63 is a remarkably accurate indicator of when a person was born. Age is determined to within 1.6 years, whereas the commonly used morphological evaluation of skeletal remains and tooth wear is sensitive to within 5–10 years in adults.

The amount of carbon-14 isotope (^{14}C) in the atmosphere remained relatively stable until 1955, when above-ground nuclear bomb tests caused it to rise dramatically^{1,2}. Although the bombs were detonated at only a few locations, the additional ^{14}C in the atmosphere rapidly equalized around the globe. Since the Test Ban Treaty in 1963, atmospheric ^{14}C has been dropping exponentially (Fig. 1a). This is not primarily because of radioactive decay (the half-life of ^{14}C is 5,730 years) — it is also due to diffusion from the atmosphere³. Atmospheric ^{14}C reacts with oxygen to form carbon dioxide, which is incorporated into plants by photosynthesis; by eating plants, and animals that feed on plants, the ^{14}C concentration in the human body closely parallels that in the atmosphere at any given time^{4–6}.

The enamel of individual teeth is formed at distinct, well characterized times during childhood^{7,8} and it contains 0.4% carbon. There is no turnover of enamel after it has been laid down, so the ^{14}C concentration reflects that in the atmosphere at the time of enamel formation. We measured the ^{14}C content of tooth enamel (for methods, see supplementary information) and related it to the known concentrations in the atmosphere in different years to establish the year of tooth formation. This date was then related to the known age for enamel deposition of individual teeth⁷ to establish the person's year of birth (Fig. 1a).

We found that this method gave a remarkably precise estimate of age for 22 individuals ($R^2 = 0.99$ from regression shown in Fig. 1b; for details, see supplementary information). The average systematic deviation from the correct value was $+0.2$ years, and the average absolute error for individual measurements was 1.6 ± 1.3 years (s.d.). This indicates that the precision is substantially higher than that obtained by other available methods⁹.

The final formation of enamel is for the

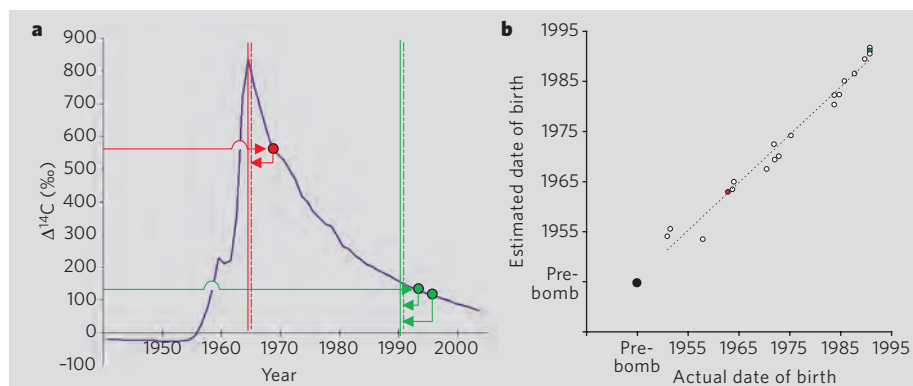


Figure 1 | Date of birth determined from ^{14}C in teeth. **a**, Nuclear bomb tests during 1955–63 produced large amounts of ^{14}C , which have since declined exponentially (blue line). The ^{14}C :C ratio has ranged from 1.15×10^{-12} to 2.20×10^{-12} since 1950. $\Delta^{14}\text{C}$ represents the ^{14}C value corrected for radioactive decay and ^{13}C fractionation (see supplementary information). To estimate an individual's date of birth, the ^{14}C concentration measured in their tooth enamel is plotted on to the curve of atmospheric $\Delta^{14}\text{C}$ against time (blue) to find the year of enamel synthesis (right-pointing arrows), and the known age at enamel formation for individual teeth was then subtracted from the year obtained to give the date of birth (left-pointing arrows; dashed vertical lines). Two representative cases are shown (red and green); two teeth were analysed for the case depicted in green. Solid vertical lines, actual dates of birth. **b**, Relation between estimated and actual dates of birth. Each point corresponds to one individual, except for the 'pre-bomb' point, which represents four individuals; coloured points are cases shown in **a**.

wisdom teeth at 12 years of age. For individuals born before 1943 (12 years before the onset of nuclear bomb testing), we can therefore conclude by this method only that birth occurred before that year, albeit with a high degree of certainty (100% correct in our analysis (Fig. 1b); $n = 4$). In any case of ambiguity as to whether birth occurred before or after the peak of nuclear bomb testing, it is necessary to analyse two teeth that were formed at different ages: this distinguishes whether the ^{14}C measurements relate to the rising or falling part of the ^{14}C curve (Fig. 1a).

The sensitivity of our method is mainly determined by variation between individuals in their age at tooth formation, and the precision of the ^{14}C measurement. The degree of inter-individual variation is different for different teeth, so selection for ^{14}C measurement of teeth with the least variation⁸ and of several teeth from the same individual should give a more accurate date of birth. With regard to measurement precision, we cannot exclude the possibility that differences in diet or in local conditions might contribute some variability in the amount of ^{14}C incorporated into tooth enamel. Although such an effect is not supported by results from comparative analyses of different foodstuffs produced in rural and industrial areas¹⁰, the method will need to be

verified on a larger number, and perhaps on a wider geographical range, of cases before it can be applied to forensic work.

Although the nuclear bomb tests were conducted several decades ago and the resulting change in atmospheric ^{14}C is now decreasing only slowly (Fig. 1a), the method described here should allow precise age determination for a long time to come because techniques for ^{14}C measurement are becoming increasingly sensitive. In addition, accelerator mass spectrometry for ^{14}C analysis has become more accessible and inexpensive, making the potential application of our dating method no more difficult than other methods now used in routine forensic examinations.

Kirsty L. Spalding*, Bruce A. Buchholz‡, Lars-Eric Bergman†, Henrik Druid†, Jonas Frisén*

Departments of *Cell and Molecular Biology, Medical Nobel Institute, and †Forensic Medicine, Karolinska Institute, 17177 Stockholm, Sweden
e-mail: jonas.frisen@cmb.ki.se

‡ Center for Accelerator Mass Spectrometry, Lawrence Livermore National Laboratory, Livermore, California 94551, USA

- De Vries, H. *Science* **128**, 250–251 (1958).
- Nydal, R. & Lovseth, K. *Nature* **206**, 1029–1031 (1965).
- Levin, I. & Kromer, B. *Radiocarbon* **46**, 1261–1272 (2004).
- Libby, W. F., Berger, R., Mead, J. F., Alexander, G. V. &

- Ross, J. F. *Science* **146**, 1170–1172 (1964).
 5. Harkness, D. D. *Nature* **240**, 302–303 (1972).
 6. Spalding, K. L., Bhardwaj, R. D., Buchholz, B. A., Druid, H. & Frisén, J. *Cell* **122**, 133–143 (2005).
 7. Nolla, C. M. *J. Dent. Child.* **27**, 254–266 (1960).
 8. Bolanos, M. V., Manrique, M. C., Bolanos, M. J. & Briones, M. T. *Foren. Sci. Int.* **110**, 97–106 (2000).
 9. Ritz-Timme, S. et al. *Int. J. Legal Med.* **113**, 129–136 (2000).

10. Otlet, R. L., Walker, A. J., Fulker, M. J. & Collins, C. *J. Envir. Radioact.* **34**, 91–101 (1997).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared (see online version of the communication).

doi:10.1038/437333a

BOTANY

Floral fluorescence effect

The way flowers appear to insects is crucial for pollination^{1–3}. Here we describe an internal light-filtering effect in the flowers of *Mirabilis jalapa*, in which the visible fluorescence emitted by one pigment, a yellow betaxanthin, is absorbed by another, a violet betacyanin, to create a contrasting fluorescent pattern on the flower's petals. This finding opens up new possibilities for pollinator perception as fluorescence has not previously been considered as a potential signal in flowers.

We investigated the spectra and distribution of the pigments in the multicoloured, strikingly patterned flowers of *M. jalapa* (Nyctaginaceae), which open only in the late afternoon. This and related plants, such as *Bougainvillea*, *Celosia*, *Gomphrena* and *Portulaca*, contain pigments known as betalains. These comprise the yellow, fluorescent betaxanthins⁴ and violet betacyanins, of which betanin (betanidin-O- β -glucoside) is the most common.

We extracted and purified the pigments of *M. jalapa* flowers and analysed them by high-performance liquid chromatography, as previously described⁵. The analysis confirmed that the pigmentation pattern on the flowers was due to a mixture of betaxanthins and betanins. Measurement of the fluorescence-emission

spectrum of dopaxanthin and the absorbance spectrum of betanin indicates that the light emitted by the fluorophore is strongly re-absorbed (Fig. 1). Addition of increasing concentrations of betanin to the dopaxanthin solution reduced the intensity of its fluorescence, until only 30% of the initial fluorescence was detectable at a ratio of 8.5:1. (For details and methods, see supplementary information.)

This internal light-filtering effect between the two types of betalain plant pigment causes a fading of visible fluorescence on parts of the flower where both types are present; areas containing only betaxanthins appear yellow under white light because of a combination of fluorescence and reflectance of non-absorbed radiation (Fig. 2a). The effect can be demonstrated in a system designed to visualize green fluorescence, which filters the incident light to blue and causes betaxanthins in the flower to fluoresce by emitting green light (Fig. 2b).

Detailed images of different zones of petal coloration were obtained by using light and fluorescence microscopy. A brightfield image under white light shows some cells containing only betaxanthins (Fig. 2c, yellow), others with betacyanins (Fig. 2c, deep-red spots), and some with both pigments together (Fig. 2c, orange).

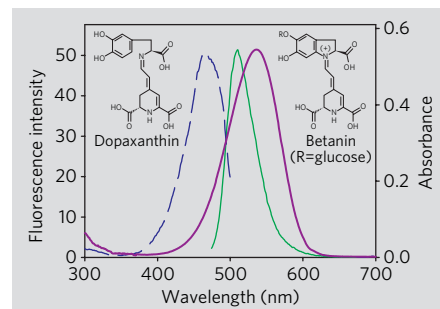


Figure 1 | Spectra of dopaxanthin and betanin.

Dopaxanthin is used as a model betaxanthin because of its structural (insets) and biochemical similarity to betacyanins. When excited by blue light, betaxanthins emit green fluorescence⁴. Fluorescence spectra (blue line, excitation spectrum; green line, emission spectrum) for natural dopaxanthin (6.0 μ M, in water) are shown; violet line, absorbance spectrum of pure betanin (8.4 μ M, in water). Note the overlap of the emission and absorbance spectra of the pigments.

The fluorescence micrograph shows that fluorescence is inhibited in areas where betaxanthins coexist with betanin (Fig. 2d) — the dark area corresponds to the orange area in Fig. 2c.

Fluorescence can be an important signal in mate choice for budgerigars⁶ and possibly in mantis shrimp⁷, and it may be that in flowers it attracts pollinators. The patterns arising from the internal light-filtering effect between betalain pigments described here could encourage bees¹ and bats⁸, which have visual receptors that are sensitive to green light and can detect bright targets better than dim ones⁹. Variation in light emission by flowers at visible wavelengths also modifies their colour, which would enhance their visibility to pollinators¹⁰.

Fernando Gandia-Herrero,

Francisco García-Carmona, Josefa Escribano

Departamento de Bioquímica y Biología Molecular A, Unidad Docente de Biología, Facultad de Veterinaria, Universidad de Murcia, 30100 Espinardo, Murcia, Spain
 e-mail: gcarmona@um.es

- Gumbert, A. *Behav. Ecol. Sociobiol.* **48**, 36–43 (2000).
- Giurfa, M., Eichmann, B. & Menzel, R. *Nature* **382**, 458–461 (1996).
- Heiling, A. M., Herberstein, M. E. & Chittka, L. *Nature* **421**, 334 (2003).
- Gandia-Herrero, F., García-Carmona, F. & Escribano, J. *J. Chromatogr. A* **1078**, 83–89 (2005).
- Gandia-Herrero, F., Escribano, J. & García-Carmona, F. *Plant Physiol.* **138**, 421–432 (2005).
- Arnold, K. E., Owens, I. P. F. & Marshall, N. J. *Science* **295**, 92 (2002).
- Mazel, C. H., Cronin, T. W., Caldwell, R. L. & Marshall, N. J. *Science* **303**, 51 (2004).
- Winter, Y., Lopez, J. & von Helversen, O. *Nature* **425**, 612–614 (2003).
- De Ibarra, N. H., Vorobyev, M., Brandt, R. & Giurfa, M. *J. Exp. Biol.* **203**, 3289–3298 (2000).
- Vorobyev, M., Marshall, J., Osorio, D., De Ibarra, N. H. & Menzel, R. *Color Res. Appl.* **26** (suppl.), 214–217 (2001).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

doi:10.1038/437333a

BRIEF COMMUNICATIONS ARISING online
 ♦ www.nature.com/bca see Nature contents.

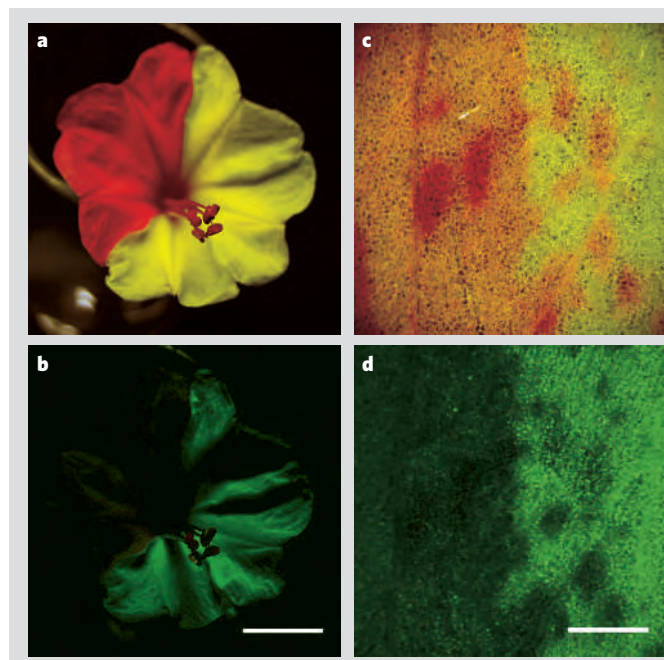


Figure 2 | Visible fluorescence in *Mirabilis jalapa* petals.

a, b, Flower with areas of red or yellow coloration under white light (**a**); only the yellow areas emit green fluorescence when excited by blue light (**b**) (scale bar, 1.5 cm). **c, d**, Light micrographs of a section of a single red-and-yellow petal, showing brightfield (**c**) and fluorescent (**d**; excitation wavelength, 450–490 nm) images (scale bar, 500 μ m). Green fluorescence is due to betaxanthins; dark areas correspond to orange areas in **c**, where light emitted from the fluorescent pigment is absorbed by betanin.

DNA POLYMERASES

Hoogsteen base-pairing in DNA replication?

Arising from: D. T. Nair, R. E. Johnson, S. Prakash, L. Prakash & A. K. Aggarwal *Nature* 430, 377–380 (2004).

Human polymerase- ϵ belongs to the error-prone Y family of polymerases, which frequently incorporate incorrect nucleotides during DNA replication but can efficiently bypass DNA lesions^{1,2}. On the basis of X-ray diffraction data, Nair *et al.* propose that Hoogsteen base-pairing is adopted by DNA during its replication by this enzyme³. Here I re-examine their X-ray data and find that the electron density is very weak for a Hoogsteen base pair formed between a template adenine deoxyribonucleotide in the *syn* conformation and a deoxythymidine 5'-triphosphate (dTTP), and that the fit is better for a normal Watson–Crick base pair. As a guanine–cytosine (G–C) base pair has no potential to form a Hoogsteen base pair at physiological pH, Hoogsteen base-pairing is unlikely to be used in replication by this polymerase.

Figure 1 of Nair *et al.*³, which describes an asymmetric complex with two polymerase molecules bound by one dTTP/DNA duplex,

is misleading because there were two such asymmetric complexes in one asymmetric unit in their coordinate set (Fig. 1a; Protein Data Bank accession number 1T3N), each present at 50% occupancy. Four polymerase and two DNA duplex molecules in the two complexes were modelled and refined independently. The two complex models, when considered individually at full occupancy, yield free *R*-factors of 34.7% and 32.0%, respectively. The authors do not explain how both complexes form one coherent lattice (their two models are added on structure factors and not on intensities, as in classical twinning). With the two models combined, the electron density is completely two-fold symmetric. The authors do not discuss poor geometry and bad clashes in their structure³.

I believe that the data of Nair *et al.*³ are better explained by the atomic model shown in Fig. 1b, which has a free *R*-factor of 28.8% (Protein Data Bank accession number 1ZET).

This model is in space group $P6_522$ and contains only one polymerase and one DNA duplex, and differs from the original models in $P6_5$ by an origin shift of $\Delta z = c/12$. Removal of the second model of Nair *et al.*³ effectively reduces the number of atoms per unique observation by almost half. The new crystallographic dyad is present in the original X-ray data. All reflections related by this dyad differ by only 3.8% to 2.3 Å resolution, which is smaller than the reported differences between all six-fold symmetry-related reflections of the data (merging *R*-factors in $P6_5$ of 5.2% to 8.3%).

In the new $P6_522$ space group, lattice contacts are formed entirely by the polymerase with the DNA duplex, which is enveloped inside the protein in two possible orientations (those shown in Fig. 1b, e). The second of the two DNA orientations in Fig. 1e is the crystallographic two-fold-symmetry mate of the orientation in Fig. 1b. The polymerase–lattice symmetry is statistically maintained owing to a completely random distribution of the two DNA orientations, with 50% occupancy in each orientation (Fig. 1e). Similar cases of 'substrate' disorder are observed with the dimeric protease of human immunodeficiency virus bound to an asymmetric inhibitor^{4,5} and with *Taq* polymerase complexed with DNA⁶. In addition, three high-resolution structures of complexes have been obtained in which asymmetric DNA duplexes are trapped in two orientations inside polymerase lattices while the crystal symmetry determined by the polymerase is statistically maintained (S. H. Eom, J. W. and T. A. Steitz, unpublished results).

Electron-density maps indicate that a normal Watson–Crick base pair fits as well as, or better than, a Hoogsteen base pair, irrespective of whether the maps are calculated using the original model with the replicating base pair omitted (Fig. 1c) or the corrected model (Fig. 1d). The accuracy of the original model suffered from model bias and overfitting problems owing to the artificial quadrupling of the number of atoms relative to the corrected model. The original cross-validation test was also invalid because of the unrecognized dyad relating 90% of the test set to the working set.

Jimin Wang

Center for Structural Biology, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA
e-mail: wang@csb.yale.edu

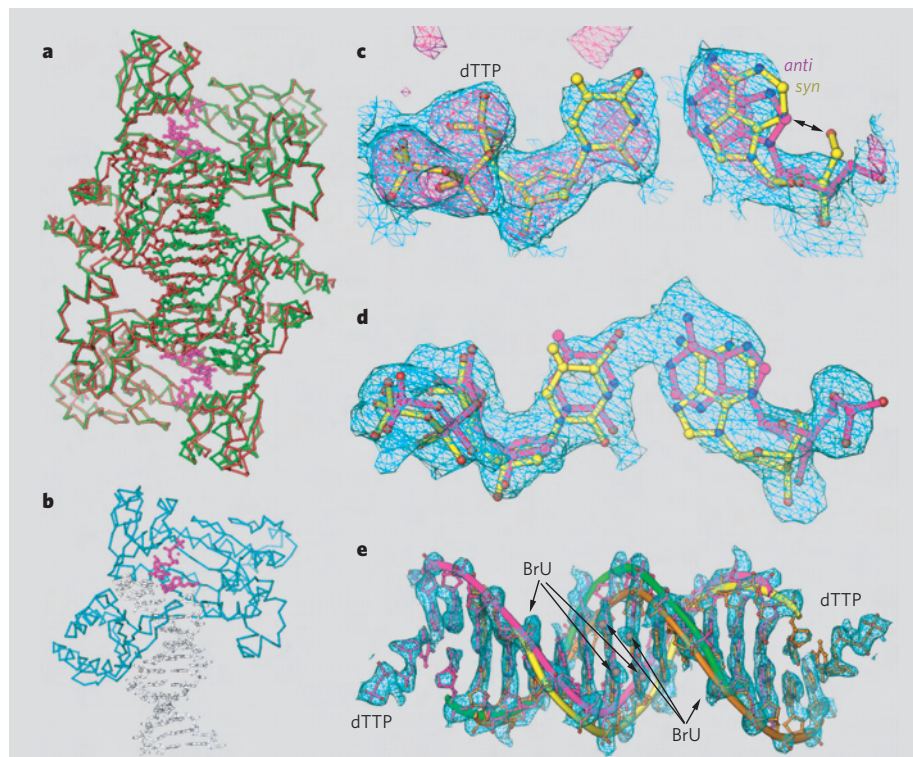


Figure 1 | A normal Watson–Crick base pair fits the electron-density map as well as, or better than, a Hoogsteen base pair in the human polymerase- ϵ /DNA complex. **a**, Two asymmetric complexes (green and red) are present in the coordinates in $P6_5$ instead of one (green; as shown in Fig. 1 of Nair *et al.*³). The replicating base pairs in question are depicted in magenta. **b**, The corrected model, with one polymerase and one DNA duplex in $P6_522$. **c**, Superposition of the replicating base pair in the Hoogsteen (yellow) or Watson–Crick (magenta) configuration, with the omitted $F_o - F_c$ (magenta, 3σ) and $2F_o - F_c$ (cyan, 0.5σ) electron-density maps calculated in $P6_5$ using the model of Nair *et al.*³ minus the replicating base pair. A double arrow indicates a steric clash between the 5'-OH and the base in the *syn* conformation. **d**, **e**, A composite-omit $2F_o - F_c$ map (cyan, 0.2σ in **d** and 1.0σ in **e**) prepared using the corrected model before the DNA duplex and dTTP were built in. BrU, 5-bromodeoxyuridine.

1. Johnson, R. E., Washington, M. T., Haracska, L., Prakash, S.

- & Prakash, L. *Nature* **406**, 1015–1019 (2000).
2. Tissier, A., McDonald, J. P., Frank, E. P. & Woodgate, R. *Genes Dev.* **14**, 1642–1650 (2000).
 3. Nair, D. T., Johnson, R. E., Prakash, S., Prakash, L. & Aggarwal, A. K. *Nature* **430**, 377–380 (2004).
 4. Dreyer, G. B. et al. *Biochemistry* **31**, 6646–6659

- (1992).
5. Murthy, K. H., Winborne, E. L., Minnich, M. D., Culp, J. S. & Debouck, C. J. *Biol. Chem.* **267**, 22770–22778 (1992).
 6. Eom, S. H., Wang, J. & Steitz, T. A. *Nature* **382**, 278–281 (1996).

doi:10.1038/nature04199

DNA POLYMERASES

Nair et al. reply

Reply to: J. Wang *Nature* **437**, doi:10.1038/nature04199 (2005)

We have solved the crystal structure of human DNA polymerase- ϵ (hPol ϵ) bound to a template primer and incoming nucleotide, and shown that hPol ϵ uses Hoogsteen base-pairing for normal DNA synthesis¹. Jimin Wang questions the space group we used and claims that Hoogsteen base-pairing is not supported by our electron-density maps². We show here that the issue of space group does not affect our conclusions¹ and that our electron-density maps are clear in indicating that Hoogsteen and not Watson–Crick base-pairing occurs in this situation.

We are aware that our data merge well with 6/*mmm* symmetry and, in fact, we calculated the initial experimental (MAD-phased) map in space group *P*6₅22. The map revealed two DNA polymerases, one at the replicative end of the template primer and the other at the blunt end. This was unexpected as there was no precedence for such a binding arrangement. Because the two ends are chemically and structurally distinct (for example, the replicative end has unpaired template bases and a bound incoming nucleotide, whereas the blunt end does not), the two polymerases could bind slightly differently, and they do. We therefore refined the structure in space group *P*6₅, rather than *P*6₅22, so as not to force an exact identity on the two polymerases.

In addition, data from a brominated derivative indicated that the complex packs in two orientations — ‘up’ in about half the unit cells and ‘down’ in the others, as occasionally occurs in crystals of protein–DNA complexes³. We therefore refined the structure in *P*6₅ with two orientations — as the more prudent approach to the packing in crystals. However, whether the structure is refined in space group *P*6₅ or *P*6₅22 has no bearing on our conclusion that replication by hPol ϵ occurs by Hoogsteen and not Watson–Crick base-pairing¹.

This conclusion is based on electron-density

maps, which also include simulated annealing omit maps — the gold standard for the evaluation of structure. For the calculation of such maps, the structure (with the template nucleotide omitted) is heated to 1,000 K and then slowly cooled to remove model bias. Figure 1 shows a portion of the simulated annealing 2*F*_o–*F*_c omit map contoured at 0.8 σ . The only way to fit this and other maps is with a Hoogsteen base pair: for a Watson–Crick base pair, either the sugar phosphate is far out of density or the bases are too close, with non-linear and short hydrogen bonds and a C1′–C1′ distance that is much smaller than the expected distance of about 10.5 Å.

It is not clear how Wang calculates his maps², but almost any base-pair arrangement could be fitted at these low σ values (compare his Fig. 1c,d at 0.5 σ and 0.2 σ , respectively, with ours contoured at 0.8 σ). Even then, the Watson–Crick base pair seems to us to be rather ill-fitted, with short and nonlinear hydrogen bonds. A single base pair makes a minor contribution to the overall scattering, and therefore can be refined in any arrangement without any significant effect on *R*-factor or *R*-free. The geometry of our structure is reasonable¹ (r.m.s.d._{bond} = 0.006 Å; r.m.s.d._{angle} = 1.3 deg) and there are no steric clashes flagged during refinement. Wang points to a distance of 2.4 Å to the O5′ atom (his Fig. 1c), but this is a ‘terminal’ atom added by the refinement program and so can be moved slightly.

Wang claims² that, as G–C cannot form a Hoogsteen base pair at physiological pH, Hoogsteen base-pairing is unlikely to be used in replication by hPol ϵ . However, we find that in the structure of Pol ϵ bound to template guanine and an incoming dCTP, Hoogsteen base-pairing does occur¹. Also, in experiments using N7-modified analogues of purine template bases, where modifications at position N7 disrupt Hoogsteen but not Watson–Crick

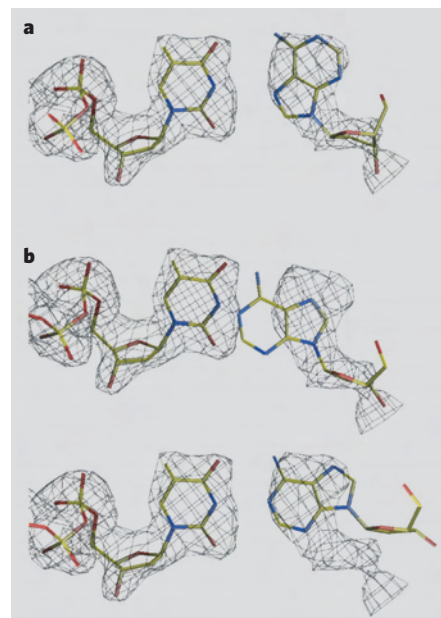


Figure 1 | Hoogsteen base-pairing (dAdTTP) in the active site of human DNA polymerase- ϵ .

A simulated annealing 2*F*_o–*F*_c omit map, contoured at 0.8 σ , is shown. The templating nucleotide was omitted and the structure heated to 1,000 K and then cooled (40 K per step).

a, b, Whereas a Hoogsteen base pair (**a**) fits the density, a Watson–Crick base pair (**b**) does not. In **b**, the bases are too close (top) or the sugar phosphate is far out of density (bottom).

base-pairing, we find strong inhibition of DNA synthesis by Pol ϵ , but not by all other DNA polymerases of the A, B or Y families⁵.

Our structural and biochemical evidence shows that Pol ϵ is exceptional among the known DNA polymerases in its dependence on Hoogsteen base-pairing for DNA synthesis.

Aneel Aggarwal*, Deepak Nair*, Robert Johnson†, Louise Prakash†, Satya Prakash†

*Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York, New York 10029, USA

e-mail: aggarwal@inka.mssm.edu

†Sealy Center for Molecular Science, University of Texas Medical Branch, Galveston, Texas 77755, USA

1. Nair, D. T., Johnson, R. E., Prakash, S., Prakash, L. & Aggarwal, A. K. *Nature* **430**, 377–380 (2004).
2. Wang, J. *Nature* **437**, doi:10.1038/nature04199 (2005).
3. Becker, S., Groner, B. & Muller, C. W. *Nature* **394**, 145–151 (1998).
4. Nair, D. T., Johnson, R. E., Prakash, L., Prakash, S. & Aggarwal, A. K. *Structure* (in the press).
5. Johnson, R. E., Prakash, L. & Prakash, S. *Proc. Natl Acad. Sci. USA* **102**, 10466–10471 (2005).

doi:10.1038/nature04200

**Cover illustration**

The *Porpita* is a genus of bright-coloured Siphonophora found floating in the warmer parts of the ocean. It has a greenhouse-like husk and many photosynthetic zooxanthellae.
(Credit: Peter Parks/Imagequestmarine.com.)

Editor, Nature

Philip Campbell

Insights Publisher

Sarah Greaves

Insights Editor

Lesley Anson

Consultant Editor

Rory Howlett

Production Editor

Maria Hodges

Senior Art Editor

Martin Harrison

Art Editor

Nik Spencer

Layouts

Marie-Claire Patin

Sponsorship

Claire Hines
Claudia Banks

Production

Sue Gray

Marketing

Robin Brown

Editorial Assistant

Laura Shaw

BIO-OCEANOGRAPHY

Two-thirds of our planet's surface is covered in water, yet the global importance of the oceans is only just becoming apparent.

We now know, for example, that marine microbes are responsible for about half of the Earth's primary productivity. They mediate many of the key biogeochemical processes that drive global nutrient cycles, and feedback mechanisms between marine ecosystems and the atmosphere have a fundamental role in regulating world climate.

Technological advances, ranging from microbial genomics to satellite remote sensing, have furthered our understanding of biological processes in the oceans. Microbial ecologists have in recent years identified levels of microbial functional diversity that were beyond their wildest dreams. One example are the 'anammox' bacteria — miniscule lifeforms capable of anaerobic ammonium oxidation, leading to the loss of previously fixed nitrogen from the ocean back to the atmosphere. Another important advance is the ability to culture such newly discovered microbes in the laboratory.

There is still much to be learned from more traditional whole-ecosystem approaches such as those that focus on patterns of species' abundance, nutrient limitation, food webs and community structure. Such knowledge should help us to understand, and predict, the effects of climate change on fragile oceanic environments such as polar marine ecosystems. There will also be surprises along the way. It is only in the past decade or so that we have come fully to appreciate the importance of viruses in the sea and the effect that they have on the carbon cycle.

We hope this Insight will impart some of the excitement of this rapidly evolving field. The study of biological oceanography may well hold the key to understanding global nutrient cycles in a time of rapid environmental change.

Jane Rees, Senior Editor

REVIEW ARTICLES

336 Genomic perspectives in microbial oceanography

E. F. DeLong & D. M. Karl

343 Molecular diversity and ecology of microbial plankton

S. J. Giovannoni & U. Stingl

349 Marine microorganisms and global nutrient cycles

K. R. Arrigo

356 Viruses in the sea

C. A. Suttle

362 Polar ocean ecosystems in a changing world

V. Smetacek & S. Nicol

nature
insight

Genomic perspectives in microbial oceanography

Edward F. DeLong¹ and David M. Karl²

The global ocean is an integrated living system where energy and matter transformations are governed by interdependent physical, chemical and biotic processes. Although the fundamentals of ocean physics and chemistry are well established, comprehensive approaches to describing and interpreting oceanic microbial diversity and processes are only now emerging. In particular, the application of genomics to problems in microbial oceanography is significantly expanding our understanding of marine microbial evolution, metabolism and ecology. Integration of these new genome-enabled insights into the broader framework of ocean science represents one of the great contemporary challenges for microbial oceanographers.

Marine ecosystems are complex and dynamic. A mechanistic understanding of the susceptibility of marine ecosystems to global environmental variability and climate change driven by greenhouse gases will require a comprehensive description of several factors. These include marine physical, chemical and biological interactions including thresholds, negative and positive feedback mechanisms and other nonlinear interactions. The fluxes of matter and energy, and the microbes that mediate them, are of central importance in the ocean, yet remain poorly understood. Detailed field studies over the past three decades have established the current 'microbial loop' hypothesis wherein microbes have a central position in the conversion of dissolved organic matter into higher trophic levels (Fig. 1). An explicit and comprehensive test of the microbial loop hypothesis, however, has not yet been achieved. In addition, the central role of microbial activities in maintaining the oxidative state of our planet, and biogeochemical cycles other than the carbon cycle, are not well captured in the current microbial loop hypothesis. Significant obstacles remain to be overcome in the measurement and modelling of ocean microbial processes.

For microbial oceanographers, a major challenge is to more accurately incorporate the details of diversity, physiology and ecology into oceanographic and biogeochemical models. For example, bacterioplankton community processes have often been modelled as being solely heterotrophic¹, understating the diverse physiologies and metabolic processes that operate *in situ*. Indeed, the importance of oxygenic photoautotrophic picoplankton, currently acknowledged to be dominant components of planktonic communities, was overlooked in early oceanographic models. In addition, planktonic microbial metabolism has sometimes been presumed to resemble that found in common laboratory strains, some of which have questionable ecological relevance. Recent advances, such as new cultivation approaches², cultivation-independent identification and enumeration strategies^{3,4}, and environmental genomics^{5,6}, are improving this situation. For example, we can now identify the genes and biochemical pathways that differentiate microbial species living in different habitats. It will soon be possible to map the metabolic and functional gene distributions of indigenous microbial species in space and time and within different oceanic provinces. The recent establishment of microbial observatories at selected sites will also

greatly aid in these efforts. The correlation between organism- and habitat-specific genomic features and other physical, chemical and biotic variables has the potential to refine our understanding of microbial and biogeochemical process in ocean systems.

All these advances — improved cultivation, environmental genomic approaches and *in situ* microbial observatories — promise to enhance our understanding of the living ocean system. Below, we provide a brief recent history of marine microbiology and outline some of the recent results from genomic technology and the challenges to be faced for integrating these new data into the larger fabric of ocean science.

A brief historical perspective

In the past 30 years, there has been a remarkable growth in understanding of marine microbiota (Fig. 2). During this time, researchers have recognized the crucial role that microbes play in ocean ecosystems. This stems in part from technical advances, such as improved epifluorescence microscope techniques and ATP-based biomass metrics⁷, that have revealed bacterioplankton standing stocks that are several orders of magnitude greater than had been estimated by viable counting techniques⁸. The use of radiotracers^{1,9} to estimate planktonic bacterial growth rates and turnover has also led to revised qualitative and quantitative models of microbial contributions to marine food webs^{10,11}. Research using quantitative autofluorescent cell counts and flow cytometry eventually led to the discovery of abundant photoautotrophic picoplankton, including *Synechococcus*^{12,13} and *Prochlorococcus*¹⁴ species, that dominate photosynthetic activities in open ocean gyres. Around the same time, hydrothermal vents and their rich macrofauna and microflora were discovered¹⁵, as well as the first bacterial isolates with an obligate growth requirement for elevated hydrostatic pressure¹⁶. The development of cultivation-independent phylogenetic surveys using ribosomal RNA (rRNA) sequencing^{3,17} and fluorescence *in situ* hybridization⁴ set the stage for contemporary environmental genomic studies. Soon after their development, rRNA-based phylogenetic survey techniques using the then-novel technique polymerase chain reaction (PCR) revealed the widespread distribution and abundance of several previously unrecognized marine microbial groups, including *Pelagibacter* (also known as SAR11)¹⁸, abun-

¹Department of Civil and Environmental Engineering & Division of Biological Engineering, 48-427 Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; ²School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii 96822, USA.

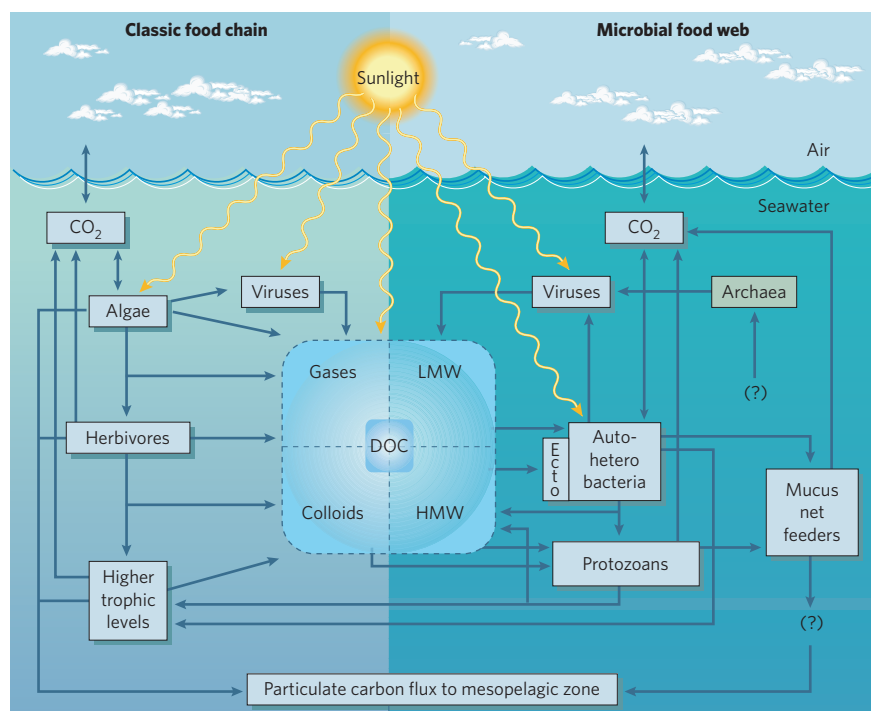


Figure 1 | Marine microbial interactions in the upper ocean. Schematic representation of the ocean food web showing on the left the classic pathway of carbon and energy flow through photosynthetic Eukarya, to herbivores and on to higher trophic levels. Depicted on the right is the microbial food web, which uses energy stored in the non-living, detrital carbon pool to produce microbial biomass that can re-enter the classic pathway of carbon and energy flow. Cell-associated ectoenzymes (Ecto) enable bacteria to use high-molecular-weight (HMW) DOC in addition to the more traditional low-molecular-weight (LMW) and gaseous carbon substances. Also shown in the microbial food web are viral particles and Archaea. At the present time, there is only rudimentary knowledge of the role of Archaea in the oceanic food web. Shown at the bottom of this diagram is the downward flux of particulate carbon (and energy), which is now thought to fuel most subeuphotic zone processes. The classic algae-herbivore grazer pathway (left side) is most important in this regard. Adapted from ref. 51.

dant new groups of planktonic marine Archaea^{19–21} and novel eukaryotic picoplankton²², to list a few examples.

More recently, the prevalence of bacteriochlorophyll-containing²³ and rhodopsin-containing²⁴ bacterioplankton was recognized, providing new perspectives on the nature of light-use strategies in ocean surface waters. Surprisingly, it has only very recently been realized that viral particles can exceed total microbial cell numbers by an order of magnitude in marine plankton²⁵ and that they represent potentially important vectors of bacterioplankton mortality and lateral gene transfer²⁶.

Many new marine microbial species continue to be described, and unexpected physiological and biochemical properties await further discovery and description. There remains much to learn about the distributions, variability and biogeochemical influence of naturally occurring microbes in the sea. To this end, new technologies are now providing life science and ocean science with remarkable new research opportunities. In particular, advances in genome science and technology are revealing the interdependences that link evolutionary, ecological and biogeochemical processes in natural marine microbial communities. For example, the co-evolution and genetic exchange between nuclear and chloroplast genomes record events that have profoundly altered the predominant modes and mechanisms of oxygenic photosynthesis on Earth²⁷.

Our understanding of the frequency and mechanisms by which genes and functional traits are acquired is being radically altered by genomic studies suggesting the pervasiveness of lateral gene transfer²⁸. Similarly, the recent recognition of the tremendous sympatric species diversity in microbial plankton is adjusting current views of form, function and speciation in the sea^{29,30}. As more samples of naturally occurring microbial genomic diversity become available, they will reveal more of the specific details of microbial evolutionary process and ecological dynamics, and how these relate to environmental processes. With the above observations and goals in mind, considerable effort is now being focused on sequencing whole marine microbial genomes as well as surveying entire marine microbial assemblages at the genomic level. The results of these efforts have the potential to improve descriptions of the structure and function of the living ocean system.

Genomics of cultivated marine microbial species

Adaptation to the unique characteristics of their environment defines the essence of marine microbial species. High salinity, low and variable nutrient concentrations and steep gradients of temperature, light, pressure and nutrients with increasing depth are some of the environmental variables that influence the evolution of form and function in microbial plankton. The adaptive instructions for coping with the marine environment are encoded in the genomic blueprints of ocean microbiota. Genome sequences are therefore useful tools for exploring the relationship between genotype, phenotype and environment in native marine microbial species (Table 1). A few examples provided below illustrate how recent sequence information has contributed to our understanding of the adaptations, physiology and ecology of marine microbes.

Genomic perspective on the forests of the sea

Diatoms, a type of algae, are one of the most conspicuous plant forms in the ocean. As a group they account for a large proportion of contemporary marine primary productivity, particularly in coastal regions. What can be learned about the remarkable evolutionary success of these protists by studying the structure of their chloroplast and nuclear chromosomes? How do genomic features of diatoms correlate with their phenotypic and ecological properties? Can genomic analyses provide new information about the metabolic and physiological properties of these important phytoplankters? Information gained from the recent sequencing of the 34-mega-base-pair (Mbp) nuclear genome of *Thalassiosira pseudonana* has answered some of these questions³¹.

The 24 diploid chromosomes of *T. pseudonana* encode genes for proteins associated with the construction of the siliceous cell wall, including genes for silicic-acid uptake, polyamine synthesis and silafin production. These newly identified genes can now be used in experiments to understand and model diatom cell-wall synthesis, and to model regulatory response mechanisms involved in growth limitation caused by silica starvation. *T. pseudonana* and other diatom species extrude chitin fibres from cell-wall pores, presumably to increase drag and thereby slow down their sinking rate. The genome of *T. pseudonana* encodes many genes for chitin synthesis and degradation, including 22 putative chitinases, which indicates a dynamic

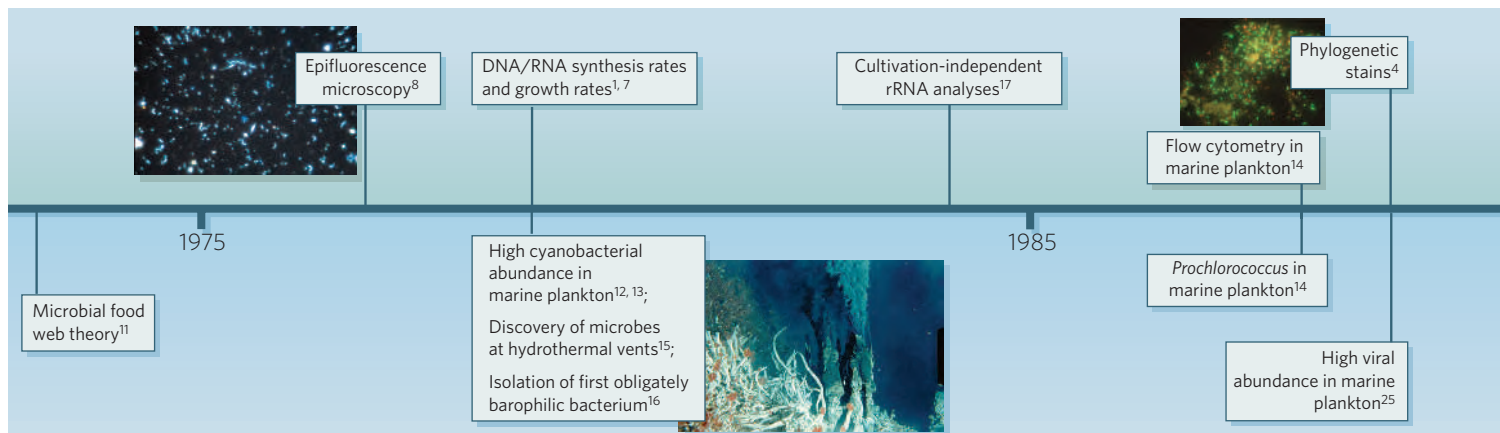
and perhaps highly regulated chitin synthesis/degradation cycle to control buoyancy and nutrient cycling³¹. Diatoms are typically growth-limited by the lack of inorganic nutrients (nitrogen, phosphorus and silica) and need to respond to the variable nutrient flux in their surrounding environment. Features of the *T. pseudonana* genome that reflect its photosynthetic, but auxotrophic, lifestyle include a wide variety of nutrient transporters, nitrogen-acquisition genes that allow conversions from several different organic and inorganic nitrogen sources and multiple pathways for energy storage using both carbohydrate and lipid synthesis degradation pathways³¹.

A genomic glimpse into coastal bacterial lifestyles

The physical, chemical and biological changes that occur along a transect from the coast to the open sea are profound. Do specific biological properties of coastal bacterioplankton differentiate them from their open-ocean relatives? Can the genomic and physiological properties of bacterioplankton explain in part their observed distribution? Can these different biological features tell us about potential regional differences in the microbial cycling of matter and energy? Although it is too early to generalize extensively, recent studies are starting to answer these questions. For example, some bacterial groups such as α -Proteobacteria in the *Roseobacter* clade seem more abundant in nearshore waters than the open ocean. The genome of *Silicibacter pomeroyi*, a member of the *Roseobacter* clade, has some adaptations that may attune this bacterium to life in coastal environments³². On the basis of its genome content, *S. pomeroyi* seems to consume not only dissolved organic carbon but also reduced inorganic compounds such as carbon monoxide and sulphur; it harbours the genes required to generate energy from these compounds. Preliminary experiments confirmed that carbon monoxide and reduced sulphur compounds are important substrates for *S. pomeroyi* metabolism³². In common with *T. pseudonana*, this coastal bacterium is equipped with a wide variety of nutrient transporters for uptake of different nitrogenous compounds such as amino acids, ammonium and urea.

The *S. pomeroyi* genome also encodes several transporters for uptake of osmolytes such as dimethylsulphoniopropionate (DMSP) and glycine betaine, presumably produced by its algal co-inhabitants, which are abundant in coastal environments³². Genomic and physiological analyses suggest a 'lithoheterotrophic' metabolic strategy for *S. pomeroyi*, allowing energy generation from organic and inorganic compounds alike. On the whole, these genomic and physiological features of *S. pomeroyi* seem to be specific adaptations to microniches typical of coastal ecosystems.

Figure 2 | Selected milestones in marine microbial oceanography. This timeline highlights a few of the advances and discoveries that have influenced marine microbiology over the past 30 years. Many important contributions could not be included simply owing to space limitation. WGS, whole-genome shotgun sequencing. PCR applications have led to more discoveries of novel microbial groups not shown here. QPCR, quantitative PCR.



Variety is the spice of ocean life

Microbiologists find it convenient to organize microbial groups into functional guilds involved in specific processes, for instance sulphate-reducers, methanogens and oxygenic photoautotrophs. It is now well recognized that within any such group, or indeed even within a species, tremendous biological diversity exists. Closely related physiological and genetic variants that seem to be 'tuned' to specific ecological conditions have been referred to as 'ecotypes'³³. What might genomic blueprints tell us about the ecophysiological differences among ecotypes? Could gene-linkage analyses of such ecotypes provide information about metabolic pathway interdependence? And at the genomic fine scale, how many differences must accumulate before one ecotype is ecologically distinguishable from another?

One of the best-documented examples of environmentally tuned ecotypes are photoautotrophic *Prochlorococcus* strains isolated from different depths. The genome sequences of these ecotypes reveal some of the unique adaptive strategies employed in different parts of the depth continuum. Several high- and low-light adapted *Prochlorococcus* strains were recently sequenced^{33,34}. The low-light-adapted strain MIT9313 isolated from the bottom of the euphotic zone has significantly more genes on its 2.4 Mbp genome than its high-light-adapted relative MED4, with a genome size of 1.66 Mbp. Genome size, however, is not a strict indicator of light adaptation, because the genome of another low-light-adapted strain, *Prochlorococcus* SS120, is 1.75 Mbp^{33,34}. Genes found in the low-light strain but not in the high-light-adapted relative included those for nitrite transport and assimilation. This is consistent with the types of nutrient available at the base of the euphotic zone. Strain MED4 encoded many more genes for high-light-inducible proteins than its low-light-adapted cousin. By contrast, the low-light-adapted strain MIT9313 contained more genes associated with photosynthesis.

The genome sequence of *Synechococcus* WH8102 revealed different characteristics from those found in its *Prochlorococcus* cousins. The *Synechococcus* genome suggests that it is more versatile generalist than *Prochlorococcus*, with a broad range of nutrient acquisition and catabolic capabilities³⁵. All told, the genomic characteristics of these different marine oxygenic photoautotrophs are consistent with, and predictive of, their ecophysiology and environmental distributions.

Expanding databases

Whole-genome sequencing of marine microbes from different habitats and phylogenetic lineages promises to widen the scope of future comparative genomic analyses^{36–38}. Other recently published marine microbial genome sequences include those of the psychrophilic (cold-loving) sulphate-reducer *Desulfotalea psychrophila*³⁸, a bioluminescent squid symbiont *Vibrio fischeri*³⁹, the marine plantomycete *Rhodopirellula baltica*³⁶, a heterotrophic bacterium from a hydrothermal vent site *Idiomarina loihiensis*³⁷ and the pressure-adapted deep-sea bacterium *Photobacterium profundum*⁴⁰. Many other marine microbes are in the pipeline for sequencing. The Gordon and Betty Moore

Foundation is currently funding a marine microbiology initiative that in part aims to generate draft genome sequences from many ecologically relevant marine microbial species (http://www.moore.org/program_areas/science/initiatives/marine_microbiology/initiative_marine_microbiology.asp).

The continued comparative analyses of marine microbial genomes promises to enhance both the databases and our current understanding. Nevertheless, the phenomenal 'microdiversity', disparate evolutionary histories and variable ecological strategies of marine microbial species necessitates careful sampling of more habitats and acquisition of even more genome sequences, from diverse as well as closely related species. New and improved methods for cultivating commonly occurring, but previously cultivation-resistant, microbial species are now being successfully developed², and these should greatly increase the relevance and use of existing genomic databases. An excellent example of this was the cultivation of one of the most abundant bacteria in the ocean *Pelagibacter ubique*^{2,41}. The genome sequence of this ubiquitous marine microbe revealed its streamlined nature, as well as the presence of a proteorhodopsin-based photophysiology^{24,41}. Continuing efforts will help to provide a comparative database for assessing the genomic repertoire of cultivated microbial species as well as the crucial context necessary for interpreting nascent environmental genomic surveys of marine microbial communities (see below).

Genomic surveys of natural microbial communities

Biological processes occurring in natural microbial communities have diverse, complex, interdependent intracellular and intercellular reactions. Much of this complexity is encoded in the structures, distribution and dynamics of interacting genomes in the environment. Recently, it has become possible to directly access the genomes of co-existing microbial species in natural communities en masse, without cultivation, using environmental genomic approaches^{6,29}. Various terms environmental genomics, metagenomics or ecogenomics, cultivation-independent genomic approaches can provide a new perspective on the naturally occurring microbial world.

Some of the first environmental genomic studies were focused on marine picoplankton^{41,42}. Initially, large genomic fragments were cloned directly from marine microbial communities to survey genomic features of then-uncultivated phylogenetic groups, such as marine Archaea^{5,24,42}. Similar genomic surveys recently led to the discovery of rhodopsins in bacteria, a domain of life not anticipated to contain these photoproteins²⁴. Hypotheses about these novel bacterial rhodopsins could be tested *in vitro* and *in vivo* by heterologous expression in *Escherichia coli* and subsequent biophysical and biochemical characterization²⁴. Subsequent field studies revealed that the bacterioplankton rhodopsins were diverse, widespread and abundant in the marine environment^{43,44}. The progression from genome sequence to biochemical and biophysical experimentation to more refined ecological observations presents a promising model for future work aimed at obtaining new genome sequence data for physiological, ecological and

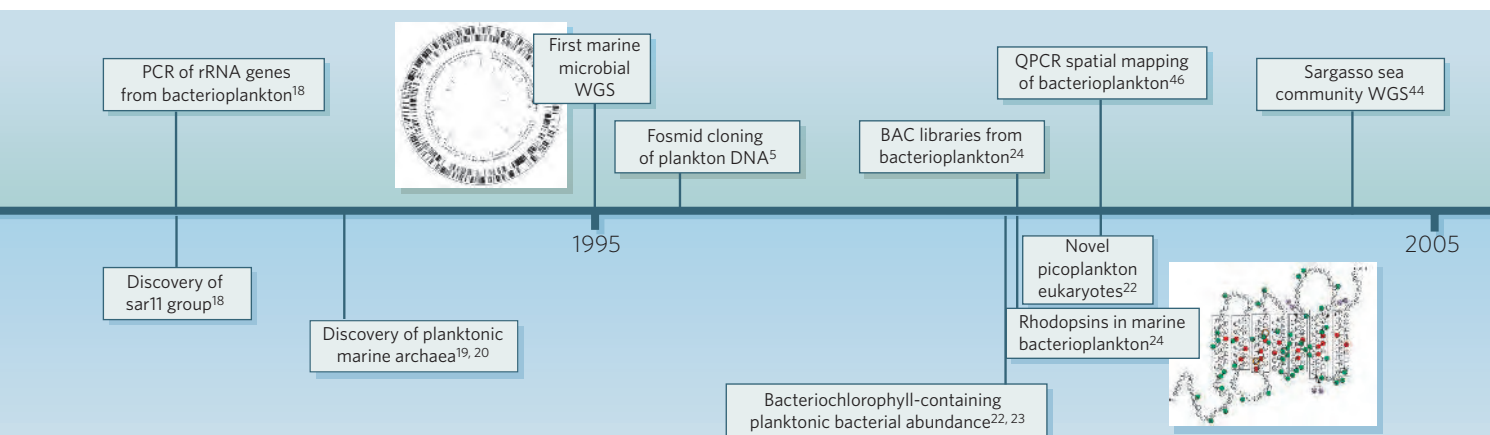
oceanographic research²⁹.

Another recent field application of marine genomics employed whole-genome shotgun (WGS) sequencing approaches. Spectacularly successful for sequencing whole microbial genomes, WGS involves the random sequencing of small (~3 kbp) DNA insert libraries and subsequent assembly of long consensus DNA sequences *in silico*. The high-throughput nature of this approach makes it extremely attractive, but environmental variation in microbial species richness, evenness and intraspecific genetic heterogeneity pose challenges to reassembling large contiguous DNA sequences (contigs) from smaller sequence fragments²⁹. Nevertheless, large amounts of data can be recovered by WGS approaches. The most marked example was the recent genomic survey of the Sargasso Sea microbial assemblages, where over 1.6 billion bp of genome-sequence information and about 1.2 million genes were recovered from the collective microbial assemblage⁴⁴.

The raw power of a massive WGS DNA-sequencing survey was revealed in the Sargasso Sea study²⁹. A remarkable number of new genes and gene families were recovered in this data set, and gene-mining efforts will continue to yield useful results. The reliable extraction of contiguous, long DNA-sequence assemblies, however, from short DNA sequences derived from complex, heterogeneous populations still pose a challenge to current assembly methods and algorithms. It seems evident from ongoing analyses that the availability of appropriate and relevant reference genomes, as well as the availability of large insert clone sequences as a reference, will play an important role in interpreting WGS sequencing surveys from complex microbial communities²⁹.

New approaches to the interpretation and analysis of microbial assemblage WGS data sets continue to be developed. Tringe *et al.*⁴⁵ recently compared, on a gene-by-gene basis, the similarities and differences between microbial communities from the Sargasso Sea, a whalebone, an acid-mine drainage assemblage and a farm silage assemblage using WGS data sets⁴⁵. By taking a 'gene-centric', as opposed to an assembly-driven 'genome-centric', approach, it was possible to compare the patterns of occurrence of specific gene categories and assemble 'community profiles' of functional gene content. The over-represented specific categories of 'environmental gene tags' (EGTs) in different samples (for instance, a disproportionate representation of photosynthetic genes and rhodopsins in the Sargasso Sea sample) verified the usefulness of this approach for identifying metabolic features unique to specific microbial assemblages⁴⁵.

There are several ways in which microbial-community genomic data are likely to be applied in microbial oceanography and ecology in the future. First, a comprehensive microbial 'parts list' (for example, gene, taxon and biochemical pathway content) of the ocean's microbial inhabitants will provide a deeper perspective of their physiological potential. These data help formulate biochemical and physiological hypotheses that can be tested in the lab and in the field. Second, detailed sequence-based genomic data sets will also provide informatic tools and reagents, in the form of molecular probes and predicted mol-



ecular and organic geochemical targets, to better track and quantify specific microorganisms and microbial processes. For example, it is now becoming possible to quantitatively map multiple microbial groups in time and space using quantitative PCR⁴⁶, and in the future DNA-microarray techniques are likely to play a significant role as well. Improved gene and gene-expression mapping and inventories in the field, as well as more targeted and specific quantitative techniques, may lead to more meaningful biogeochemical measurements. This quantitative gene, taxon and metabolic mapping may, in turn, enable the development of models that more accurately reflect the *in situ* biological details. Third, comparative analyses of entire microbial assemblages have the potential to reveal larger-scale patterns of biochemical interactions and habitat-specific correlations that might otherwise be missed in autoecological studies of individual species. For example, comparative community genomics can reveal previously unrecognized biological interactions. For instance, rhodopsin-mediated light utilization in bacteria was entirely unknown until environmental genomics was used²⁴.

Ultimately, it may be possible to overlay microbial functional genes, metabolic pathways and taxonomic distributions upon other physical, chemical and biological oceanographic data to map microbial features onto oceanic provinces. Finally, the study of genome structure and dynamics *in situ* will provide the necessary information to interpret evolutionary processes — such as genetic drift, recombination and lateral gene transfer — that drive microbial adaptation and divergence. These data in turn provide predictive tools to model the variable complexity of form, function and process that underpin the ocean's physical, chemical and biotic interactions⁴⁷.

Contemporary challenges in microbial oceanography

One common criticism of current massive data-collection efforts is that much information, but little knowledge, is accumulating. Although it is true that the informatics challenges are significant, they are not insurmountable. The immediate tasks ahead are to organize, standardize and coordinate annotation, access and integration of environmentally derived genomic data sets. The research community needs to recognize that environmental genomic data sets are fundamentally different from other existing databases because they are not derived from an individual strain, species, taxon or a single genetic locus. Rather, each discrete data set represents the collective genes recovered randomly from one single microbial assemblage. These data sets derive from dynamic microbial populations and environments, where species richness, evenness and composition (and so gene richness, evenness and composition) are variable in space and time. This is one major reason why any attempts at assembly need to be qualified and specified, because individual sequences in any given assembly could be derived from different samples, times and origins⁴⁴. The original sequence reads, and their metadata, need to be linked explicitly with assemblies because they are of potentially of heterogeneous origin.

Many other informatic challenges need to be considered as the field of environmental genomics develops. Management, annotation and archiving of environmental data sets is an ongoing process and many, but not all, of the mechanisms to handle the large datasets now emerging are available. One issue of concern is establishment of standards and requirements for environmental genomic metadata. Additionally, tools and algorithms for the annotation of environmental genomic data sets need to be better developed. Hidden Markov model methods, effective for gene-calling in whole-genome sequences, are not as

Table 1 | Phenotypic and genomic characteristics of some recently sequenced marine microbes and communities

Isolate or environmental gene survey	Habitat or origin/habitat	Environmentally relevant phenotypic and genomic features	References
<i>Thalassiosira pseudonana</i> (K)	Surface water, plankton	Silicic acid metabolism, chitin buoyancy mechanisms, diverse nitrogen acquisition and metabolism, several energy storage and utilization strategies	31
<i>Silicibacter pomeroyi</i> DSS-3 (B)	Coastal water, plankton	Lithotrophic energy metabolism (carbon monoxide, reduced sulphur compounds), DMSP metabolism, many amino and carboxylic acid transporters	32
<i>Prochlorococcus marinus</i> MED4, SS120 (B)	Oligotrophic water, upper photic zone	High-light adapted, greater number of high-light-inducible proteins, photolyase genes, narrower range of nitrogen sources (nitrite reductase and transporter absent), fewer chlorophyll-binding antenna protein genes	33,34
<i>Prochlorococcus marinus</i> MIT9313 (B)	Oligotrophic waters, lower photic zone	Low-light adapted, smaller number of high-light-inducible proteins, no photolyase, wider range of nitrogen sources (nitrite reductase and transporter present), greater number of chlorophyll-binding antenna protein genes	33
<i>Synechococcus</i> sp. WH8102 (B)	Open ocean surface water	Wide range of nitrogen sources (organic, nitrate, nitrite), amino-acid transporters, swimming motility, more transporters than <i>Prochlorococcus</i>	35
<i>Photobacterium profundum</i> SS9 (B)	Deep-sea amphipod	Growth pressure optima > 100 atm, polyunsaturated fatty acids in membrane lipids, differential gene expression at high and low pressure growth, trimethylamine oxide respiration, possible anaerobic amino acid fermentation	40
<i>Vibrio fischeri</i> ES114 (B)	<i>Euprymna scolopes</i> light organ	Bioluminescence, specific association with squid <i>Euprymna scolopes</i> , many adhesions and toxin-like genes shared in common with other vibrio pathogens	39
<i>Rhodopirellula baltica</i> (B)	Kiel fjord seawater	Derived absence of peptidoglycan, formaldehyde oxidation pathway, abundant sulphatases, high proportion of genes most similar to Eukarya	36
<i>Desulfotalea psychrophila</i> Lsv54(B)	Permanently cold Arctic marine sediment	Growth below 0 °C, sulphate-reducer, potential oxygen-reducing cytochromes	38
<i>Idiomarina loihiensis</i> (B)	Mixed hydrothermal vent/cold seawater (1,296 m)	Amino-acid fermentation, incomplete amino acid biosynthetic pathways, exopolysaccharide and biofilm formation, reduced carbohydrate metabolism	37
<i>Pelagibacter ubique</i>	Oligotrophic coastal and open ocean	Represents a group that makes up 25% of planktonic marine bacteria, smallest known genome of free-living organism, contains proteorhodopsin	41
Sargasso Sea environmental gene survey	Oligotrophic surface water	High species diversity, high representation of photosynthetic and rhodopsin genes, significant intraspecific microheterogeneity, poor assembly	44
Eel River Basin Archaeal methane oxidizer environmental gene	Deep-sea methane seep (550 m)	High representation of methanogen-like genes, high representation of two major Archaeal methane oxidizer phylotypes, absence of one 'universal' methanogenic gene survey	50
Whale fall microbial environmental gene survey	Deep-sea whale bone	Low diversity, many glycine-betaine transporters and type IV secretion system	45

This is an illustrative but not exhaustive list of marine microbial strains whose genomes have been fully sequenced or partial shotgun environmental gene survey efforts.

* Domain affiliation in parentheses. K, Eukarya; B, Bacteria; A, Archaea.

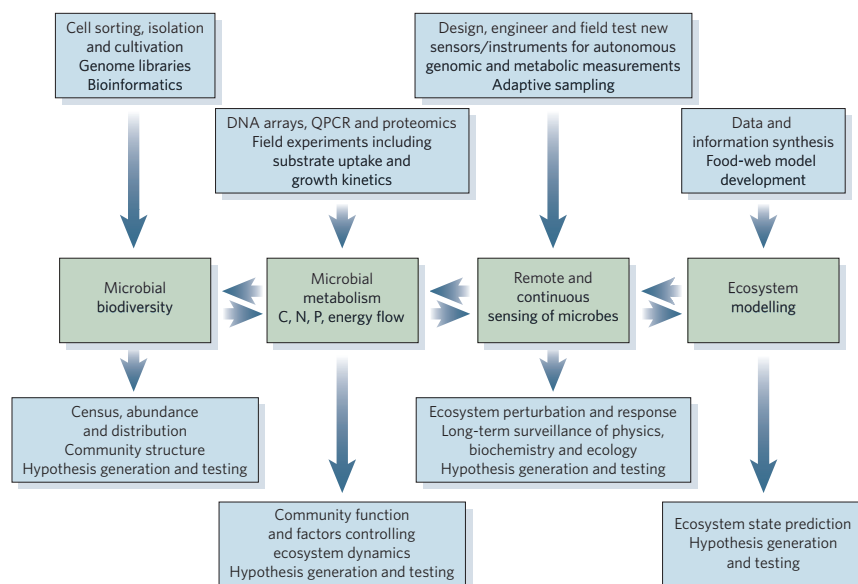


Figure 3 | Microbial systems analysis in oceanography. This flow chart outlines the research prospectus for a new international Center for Microbial Oceanography: Research and Education (CMORE), which the authors will direct. The flow path illustrates some of the linkages that will need to be fostered between and among subdisciplines in microbiology, oceanography, ecology and engineering to successfully complete the prospectus.

applicable for annotation of shorter sequences typical of metagenomic data sets. New gene-calling algorithms are another important area for development. In tandem with database issues, there is a strong need for further development of computational tools to analyse and compare microbial-community genomic data sets. Finally, interoperability between genome and metagenome sequence databases and environmental and ecological databases will be a crucial requirement in the future.

According to a recent report by the American Academy of Microbiology and by our own independent assessments, the fields of microbial ecology and genomics are currently poised at a 'crossroads of opportunity'⁴⁸. It is important to understand ocean processes and to place the emergent genomic-based knowledge of marine microorganisms into the ecological context of carbon, nitrogen, phosphorus and related biogeochemical cycles. The scientific and societal benefits at stake are profound. This research prospectus must begin with the complete integration of the disciplines of microbiology, systems biology, ecosystem dynamics and oceanography, with the goal of developing comprehensive, mechanistic-based numerical models that are essential for scaling the interactions of specific microorganisms and processes across entire ocean basins⁴⁷. Some of the tools, approaches and information flow required in this new synthesis are shown in Fig. 3. As an example of the magnitude of the challenge, consider that future climate predictions indicate a warmer and more stratified global ocean with more acidic near-surface waters, all resulting in large part from the increasing burden of carbon dioxide in the atmosphere. How will the present day oceanic 'genotype' respond to these new state variables, and are we prepared to observe and interpret the ocean's expressed 'phenotypic response'? New and integrative approaches to link microbial and global geochemical processes are required.

The establishment of a global network of ocean time-series stations for repeated observations of microbial diversity and function represents one strategy of ecosystem surveillance that would increase our understanding of these complex processes. Despite their recognized importance in Earth and ocean sciences, systematic long-term microbial observatories in key oceanic habitats are rare. Repeat observations of the North Pacific subtropical gyre over the past four decades have documented fundamental changes in microbial community structure and major bioelemental cycling. It has been hypothesized that climate variability in the North Pacific basin has led to an enhancement of microbial-based nitrogen fixation and a fundamental shift from nitrogen limitation to phosphorus limitation with several ecological consequences, including alterations in food web structure and changes in carbon sequestration⁴⁹. These changes occurred because of subtle

changes in near-surface ocean stratification and decreases in the flux of inorganic nutrients from below. These habitat changes are controlled, in part, by the frequency and duration of the basin-scale El Niño and Pacific Decadal Oscillation climate systems. The nitrogen-fixation potential of the global ocean may ultimately be controlled by the atmospheric deposition of iron, which itself is a time-variable, climate-sensitive process. Total atmospheric dust transport is affected by humankind, including population demographics, global economies and land-use patterns. These complex natural and anthropogenic forces, each with multiple potential feedback loops and possible threshold effects, have significant global impacts on microbes in the sea that are only now being appreciated. The influence of climate variability, especially that which takes place on decadal to centennial time scales, on marine microbial biogeochemistry is not well understood.

Conclusion

Marine microbes are present at billions of cells per litre in seawater. They double every few days and are consumed at about the same rate by viral parasites and protistan predators. These activities capture and process energy and drive major elemental cycles. Hidden within these dynamic assemblages and diverse genomic structures are fundamental but, at present, incomplete lessons about environmental sensing, response and adaptation, gene regulation, species and community interactions and genomic plasticity and evolution. The genomic diversity, evolutionary dynamics and ecological processes contained in these populations have global effects on the rates and flux of energy and matter in the sea, biogeochemical cycling, the Earth's atmospheric composition and global climate trends. Over the 3.8 billion years of life on Earth, microbes have been the stewards of geochemical balance and as biotic recorders of evolutionary history their 'biological memories' extend backwards further than other lifeforms. Additionally, they are Nature's biosensors *par excellence*, but we still need to decipher their code to interpret their outputs and understand the underlying strategies and mechanisms of their survival in nature. The current convergence of microbiology, ecology, genomics and ocean science has the potential to be focused in unprecedented ways through the lens of microbial oceanography. By working together, molecular biologists, microbiologists and oceanographers have new opportunities to advance observation, method and theory, which together will better describe the living ocean system. ■

1. Fuhrman, J. A. & Azam, F. Thymidine incorporation as a measure of heterotrophic bacterioplankton production in marine surface waters. *Mar. Biol.* **66**, 109–120 (1982).
2. Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**, 630–633 (2002).

3. Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R. & Stahl, D. A. Microbial ecology and evolution: a ribosomal RNA approach. *Annu. Rev. Microbiol.* **40**, 337–365 (1986).
4. DeLong, E. F., Wickham, G. S. & Pace, N. R. Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. *Science* **243**, 1360–1363 (1989).
5. Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H. & DeLong, E. F. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J. Bacteriol.* **178**, 591–599 (1996).
6. Handelsman, J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68**, 669–685 (2004).
7. Karl, D. M. Cellular nucleotide measurements and applications in microbial ecology. *Microbiol. Rev.* **44**, 739–796 (1980).
8. Hobbie, J. E., Daley, R. J. & Jasper, S. Use of Nuclepore™ filters for counting bacteria by fluorescence microscopy. *Appl. Environ. Microbiol.* **33**, 1225–1228 (1977).
9. Karl, D. M. Measurement of microbial activity and growth in the ocean by rates of stable ribonucleic acid synthesis. *Appl. Environ. Microbiol.* **38**, 850–860 (1979).
10. Azam, F. *et al.* The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* **10**, 257–263 (1983).
11. Pomeroy, L. R. The ocean's food web, a changing paradigm. *BioScience* **24**, 499–504 (1974).
12. Waterbury, J. B., Watson, S. W., Guillard, R. L. & Brand, L. E. Widespread occurrence of a unicellular marine, planktonic, cyanobacterium. *Nature* **277**, 293–294 (1979).
13. Johnson, P. W. & Sieburth, J. M. Chroococcoid cyanobacteria in the sea: a ubiquitous and diverse phototrophic biomass. *Limnol. Oceanogr.* **24**, 928–935 (1979).
14. Chisholm, S. W. *et al.* A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334**, 340–343 (1988).
15. Corliss, J. B. *et al.* Submarine thermal springs on the Galapagos Rift. *Science* **203**, 1073–1083 (1979).
16. Yayanos, A. A., Dietz, A. S. & Van Bostel, R. Obligately barophilic bacterium from the Mariana Trench. *Proc. Natl Acad. Sci. USA* **78**, 5212–5215 (1981).
17. Stahl, D. A., Lane, D. J., Olsen, G. J. & Pace, N. R. Analysis of hydrothermal vent-associated symbionts by ribosomal RNA sequences. *Science* **224**, 409–411 (1984).
18. Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. & Field, K. G. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**, 60–63 (1990).
19. DeLong, E. F. Archaea in coastal marine environments. *Proc. Natl Acad. Sci. USA* **89**, 5685–5689 (1992).
20. Karner, M. B., DeLong, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507–510 (2001).
21. Fuhrman, J. A., McCallum, K. & Davis, A. A. Novel major archaeobacterial group from marine plankton. *Nature* **356**, 148–149 (1992).
22. Moon-van der Staay, S. Y., De Wachter, R. & Vaulot, D. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409**, 607–610 (2001).
23. Kolber, Z. S., Van Dover, C. L., Niederman, R. A. & Falkowski, P. G. Bacterial photosynthesis in surface waters of the open ocean. *Nature* **407**, 177–179 (2000).
24. Béjà, O. *et al.* Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906 (2000).
25. Bergh, O., Borsheim, K. Y., Bratbak, G. & Haldal, M. High abundance of viruses found in aquatic environments. *Nature* **340**, 467–468 (1989).
26. Paul, J. H., Sullivan, M. B., Segall, A. M. & Rohwer, F. Marine phage genomics. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **133**, 463–476 (2002).
27. Falkowski, P. G. *et al.* The evolution of modern eukaryotic phytoplankton. *Science* **305**, 354–360 (2004).
28. Lerat, E., Daubin, V., Ochman, H. & Moran, N. A. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* **3**, e130 (2005).
29. DeLong, E. F. Microbial community genomics in the ocean. *Nature Rev. Microbiol.* **3**, 459–469 (2005).
30. Thompson, J. R. *et al.* Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307**, 1311–1313 (2005).
31. Armbrust, E. V. *et al.* The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**, 79–86 (2004).
32. Moran, M. A. *et al.* Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* **432**, 910–913 (2004).
33. Rocap, G. *et al.* Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**, 1042–1047 (2003).
34. Dufresne, A. *et al.* Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc. Natl Acad. Sci. USA* **100**, 10020–10025 (2003).
35. Palenik, B. *et al.* The genome of a motile marine *Synechococcus*. *Nature* **424**, 1037–1042 (2003).
36. Glockner, F. O. *et al.* Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proc. Natl Acad. Sci. USA* **100**, 8298–8303 (2003).
37. Hou, S. *et al.* Genome sequence of the deep-sea gammaproteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. *Proc. Natl Acad. Sci. USA* **101**, 18036–18041 (2004).
38. Rabus, R. *et al.* The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments. *Environ. Microbiol.* **6**, 887–902 (2004).
39. Ruby, E. G. *et al.* Complete genome sequence of *Vibrio fischeri*: a symbiotic bacterium with pathogenic congeners. *Proc. Natl Acad. Sci. USA* **102**, 3004–3009 (2005).
40. Vezzi, A. *et al.* Life at depth: *Photobacterium profundum* genome sequence and expression analysis. *Science* **307**, 1459–1461 (2005).
41. Giovannoni, S. J. *et al.* Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**, 1242–1245 (2005).
42. Schmidt, T. M., DeLong, E. F. & Pace, N. R. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* **173**, 4371–4378 (1991).
43. Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M. & DeLong, E. F. Proteorhodopsin phototrophy in the ocean. *Nature* **411**, 786–789 (2001).
44. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
45. Tringe, S. G. *et al.* Comparative metagenomics of microbial communities. *Science* **308**, 554–557 (2005).
46. Suzuki, M. T., Preston, C. M., Chavez, F. & DeLong, E. F. Quantitative mapping of bacterioplankton populations in seawater: field tests across an upwelling plume in the Monterey Bay. *Aquat. Microb. Ecol.* **24**, 117–127 (2001).
47. Doney, S. C., Abbott, M. R., Cullen, J. J., Karl, D. M. & Rothstein, L. From genes to ecosystems: the ocean's new frontier. *Front. Ecol. Environ.* **2**, 457–466 (2004).
48. Stahl, D. A. & Tiedje, J. M. *Microbial Ecology and Genomics: A Crossroads of Opportunity* (American Society for Microbiology, Washington D.C., 2002).
49. Karl, D. M., Bidigare, R. R. & Letelier, R. M. Long-term changes in plankton community structure and productivity in the North Pacific Subtropical Gyre: The domain shift hypothesis. *Deep-Sea Res. II* **48**, 1449–1470 (2001).
50. Hallam, S. J. *et al.* Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305**, 1457–1462 (2004).
51. Karl, D. M. Accurate estimation of microbial loop processes and rates. *Microbiol. Ecol.* **28**, 147–150 (1994).

Acknowledgements The authors' work is supported by the NSF, the Gordon and Betty Moore Foundation, and the Department of Energy. We thank our colleagues, students and CMORE collaborators for their ideas, inspiration and enthusiasm.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to E.F.DL (delong@mit.edu).

Molecular diversity and ecology of microbial plankton

Stephen J. Giovannoni¹ & Ulrich Stingl¹

The history of microbial evolution in the oceans is probably as old as the history of life itself. In contrast to terrestrial ecosystems, microorganisms are the main form of biomass in the oceans, and form some of the largest populations on the planet. Theory predicts that selection should act more efficiently in large populations. But whether microbial plankton populations harbour organisms that are models of adaptive sophistication remains to be seen. Genome sequence data are piling up, but most of the key microbial plankton clades have no cultivated representatives, and information about their ecological activities is sparse.

Certain characteristics of the ocean environment — the prevailing low-nutrient state of the ocean surface, in particular — mean it is sometimes regarded as an extreme ecosystem. Fixed forms of nitrogen, phosphorus and iron are often at very low or undetectable levels in the ocean's circulatory gyres, which occur in about 70% of the oceans¹. Photosynthesis is the main source of metabolic energy and the basis of the food chain; ocean phytoplankton account for nearly 50% of global carbon fixation, and half of the carbon fixed into organic matter is rapidly respired by heterotrophic microorganisms. Most cells are freely suspended in the mainly oxic water column, but some attach to aggregates. In general, these cells survive either by photosynthesizing or by oxidizing dissolved organic matter (DOM) or inorganic compounds, using oxygen as an electron acceptor.

Microbial cell concentrations are typically about 10^5 cells ml⁻¹ in the ocean surface layer (0–300 m) — thymidine uptake into microbial DNA indicates average growth rates of about 0.15 divisions per day (ref. 2). Efficient nutrient recycling, in which there is intense competition for scarce resources, sustains this growth, with predation by viruses and protozoa keeping populations in check and driving high turnover rates³. Despite this competition, steady-state dissolved organic carbon (DOC) concentrations are many times higher than carbon sequestered in living microbial biomass⁴. However, the average age of the DOC pool in the deep ocean, of about 5,000 years⁵ (determined by isotopic dating), suggests that much of the DOM is refractory to degradation. Although DOM is a huge resource, rivaling atmospheric CO₂ as a carbon pool⁶, chemists have been thwarted by the complexity of DOM and have characterized it only in broad terms⁷.

The paragraphs above capture prominent features of the ocean environment, but leave out the complex patterns of physical, chemical and biological variation that drive the evolution and diversification of microorganisms. For example, members of the genus *Vibrio* — which include some of the most common planktonic bacteria that can be isolated on nutrient agar plates — readily grow anaerobically by fermentation. The life cycles of some *Vibrio* species have been shown to include anoxic stages in association with animal hosts, but the broad picture of their ecology in the oceans has barely been characterized⁸. The story is similar for most of the microbial groups described below: the phylogenetic map is detailed, but the ecological panorama is thinly sketched. New information is rapidly flowing into the field from the

cultivation of key organisms, metagenomics and ongoing biogeochemical studies. It seems very likely that the biology of the dominant microbial plankton groups will be unravelled in the years ahead.

Here we review current knowledge about marine bacterial and archaeal diversity, as inferred from phylogenies of genes recovered from the ocean water column, and consider the implications of microbial diversity for understanding the ecology of the oceans. Although we leave protists out of the discussion, many of the same issues apply to them. Some of the studies we refer to extend to the abyssal ocean, but we focus principally on the surface layer (0–300 m) — the region of highest biological activity.

Phylogenetic diversity in the ocean

Small-subunit ribosomal (RNA) genes have become universal phylogenetic markers and are the main criteria by which microbial plankton groups are identified and named⁹. Most of the marine microbial groups were first identified by sequencing rRNA genes cloned from seawater^{10–14}, and remain uncultured today. Soon after the first reports came in, it became apparent that less than 20 microbial clades accounted for most of the genes recovered¹⁵. Figure 1 is a schematic illustration of the phylogeny of these major plankton clades. The taxon names marked with asterisks represent groups for which cultured isolates are available.

The recent large-scale shotgun sequencing of seawater DNA is providing much higher resolution 16S rRNA gene phylogenies and biogeographical distributions for marine microbial plankton. Although the main purpose of Venter's *Sorcerer II* expedition is to gather whole-genome shotgun sequence (WGS) data from planktonic microorganisms¹⁶, thousands of water-column rRNA genes are part of the by-catch. The first set of collections, from the Sargasso Sea, have yielded 1,184 16S rRNA gene fragments. These data are shown in Fig. 2, organized by clade structure. Such data are a rich scientific resource for two reasons. First, they are not tainted by polymerase chain reaction (PCR) artefacts; PCR artefacts rarely interfere with the correct placement of genes in phylogenetic categories, but they are a major problem for reconstructing evolutionary patterns at the population level¹⁷. Second, the enormous number of genes provided by the *Sorcerer II* expedition is revealing the distribution patterns and abundance of microbial groups that compose only a small fraction of the

¹Department of Microbiology, Oregon State University, Corvallis, Oregon 97331, USA.

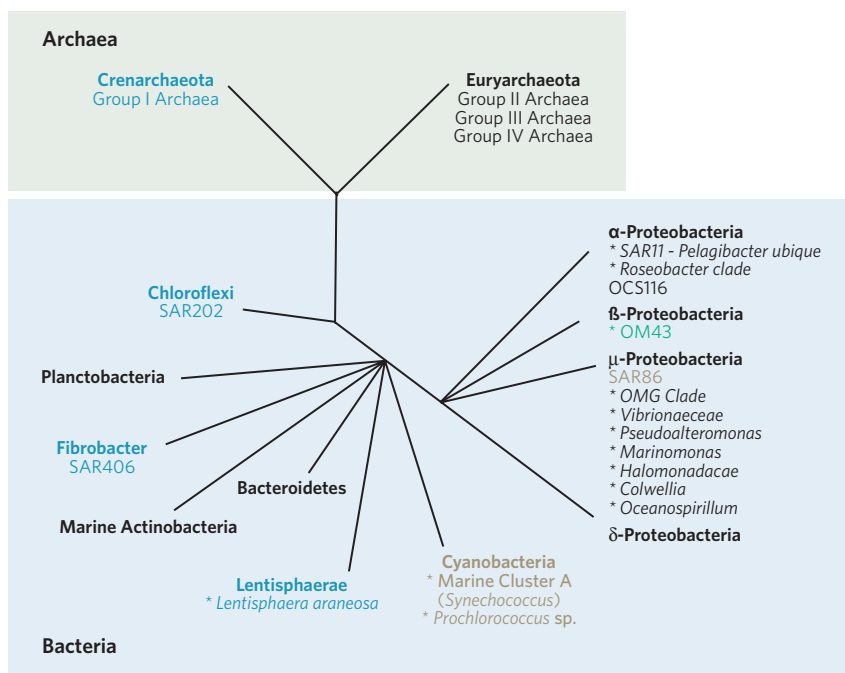


Figure 1 | Schematic illustration of the phylogeny of the major plankton clades. Black letters indicate microbial groups that seem to be ubiquitous in seawater. Gold indicates groups found in the photic zone. Blue indicates groups confined to the mesopelagic and surface waters during polar winters. Green indicates microbial groups associated with coastal ocean ecosystems.

community. As discussed below, some opportunistic strains that exploit transient conditions may fall into this category.

The interpretation of the Sargasso Sea environmental sequence data is already inspiring debate¹⁶. 16S rRNA gene sequence data from the Sargasso Sea WGS data set are shown in Fig. 2 without the sequences of *Burkholderia* and *Shewanella*, which are rare in the Sargasso Sea ecosystem and have been questioned as possible contaminants¹⁶. Naming microbial plankton by clade, as shown in Fig. 2, is a convention used by most oceanographers that is based on evolutionary principles. With few exceptions, the Sargasso Sea data fall into previously named microbial plankton clades. However, Venter has emphasized the new diversity shown by the data, concluding that 1,800 'genomic' species of bacteria and 145 new 'phylotypes' inhabited the samples recovered from the Sargasso Sea¹⁸. To reach this conclusion he applied a rule-of-thumb which assumes that 16S rRNA gene sequences that are less than 97% similar originate from different species. As we discuss further below, the origins of 16S rRNA sequence diversity within the named microbial plankton clades is a hot issue. But, however they are interpreted, the high reliability of raw WGS sequence data will be very useful for understanding the mechanisms of microbial evolution in the oceans.

Although most of the major microbial plankton clades have cosmopolitan distributions, new marine microbial plankton clades continue to emerge from studies that focus on unique hydrographic features. For example, the new Archaea Crenarchaeota and Euryarchaeota were discovered at the brine-seawater interface of the Shaban Deep, in the Red Sea¹⁹.

Patterns in time and space

Molecular biology has filled in some of the blanks about the natural history of marine microbial plankton. As genetic markers became available for ecological studies, it soon emerged that some of the dominant microbial plankton clades are vertically stratified (Fig. 1). Early indications of these patterns came from the distributions of rRNA gene clones among libraries collected from different depths^{20–23}. Although the study of microbial community stratification is far from complete, in many cases (marine unicellular cyanobacteria, SAR11, SAR202, SAR406, SAR324, group I marine Archaea) the vertical stratification of populations has been confirmed by alternative experimental approaches^{21–27}. The obvious interpretation is that many of these groups are specialized to exploit vertical patterns in physical, chemical and biological factors. A clear example is the unicellular marine

cyanobacteria. As obligate phototrophs, these cyanobacteria are confined to the photic zone. A similar pattern is found in the SAR86 clade of γ-Proteobacteria. Proteorhodopsin genes have been found in fragments of SAR86 genomes, suggesting that this clade has the potential for phototrophic metabolism.

Many of the enigmatic microbial groups for which no metabolic strategy has been identified are also stratified. The boundary between the photic zone and the dark upper mesopelagic is particularly striking — below the photic zone the abundance of picophytoplankton and SAR86 declines sharply, and marine group I Archaea, SAR202, SAR406 and SAR324 all assume a prevalent status^{21–24,26,27}. The implications of these observations are clear: the upper mesopelagic community is almost certainly specialized to harvest resources descending from the photic zone. However, with the exception of the marine group I Archaea, very little specific information is available about the individual activities of the upper-mesopelagic groups.

There are also significant differences between coastal and ocean gyre microbial plankton populations (Fig. 1; ref. 28). Typically, continental shelves are far more productive than ocean gyres because physical processes such as upwelling and mixing bring nutrients to the surface. As a result eukaryotic phytoplankton make up a larger fraction of the biomass in coastal seas, and species differ between coastal and ocean populations. Most of the bacterial groups found in gyres also occur in large numbers in coastal seas, but a number of microbial plankton clades, particular members of the β-Proteobacteria, have coastal ecotypes or appear to be predominantly confined to coastal seas²⁸.

One of the most enigmatic microbial groups in the ocean is the marine group I Archaea. Tantalizing geochemical evidence suggests that these organisms are chemoautotrophs²⁹. In the 1990s, DeLong and Fuhrman established that archaea are widely distributed and numerically significant in the marine water column^{11,20,30}. The marine group I Archaea are Crenarchaeotes. They predominantly occur in the mesopelagic, but are found at the surface in the cold waters of the southern ocean during the winter. Fluorescence *in situ* hybridization technology was used to demonstrate that marine group I Archaea populations comprise about 40% of the mesopelagic microbial community over vast expanses of the ocean, making them one of the most abundant organisms on the planet²⁴. All of the marine Archaea remain uncultured.

New data about microbial distributions has provided tantalizing hints about geochemical activity, but most progress on this question

has come either from cultures or from approaches designed to yield information from experiments performed on native populations. Fluorescence-activated cell-sorting and *in situ* hybridization have both been used to separate populations and measure their uptake of radioactive substrates^{31–33}. Testing hypotheses originating from genome sequences, oceanographers were surprised to find that the unicellular marine bacteria, particularly *Prochlorococcus*, can assimilate free amino acids — it had previously been thought that they rely solely on inorganic nitrogen³².

The species question

The question of how to name microbial plankton species is not a trivial matter. For oceanographers the issue is: where should the lines be drawn so that organisms with different properties relevant to geochemistry are given unique names? From an evolutionary perspective, the question might be phrased differently: how does one demarcate cell populations that use the same resources and possess the same suites of adaptations inherited from a common ancestor? Confusion arises from the fact that there is no general agreement about the definition of a microbial 'species'. The '97 % rule' is simple to apply but does not take into account the complex structure of microbial clades. For example, the unicellular marine cyanobacteria form a shallow clade that would constitute a single species by the 97% rule, but all agree that this clade contains several species with distinct phenotypes. Clades such as SAR86 and SAR11 are far more diverse, but can clearly be divided into subclades. One theory is that some of these 'bushy' subclades are ecotypes — populations with shared characters and unique niches³⁴.

Acinas and co-workers studied clade structure by using clone libraries prepared using PCR methods that reduced sequence artefacts¹⁷. They concluded that most sequence variation was clustered

near the tips of branches and could be attributed to neutral mutations accumulating in clonal populations. Thompson *et al.* went a step further by examining sequence divergence and genome variability in a set of 232 *Vibrio splendidus* isolates taken from the same coastal location at different times³⁵. The isolates differed by less than 1% in rRNA gene sequences, but showed extensive variation in genome size and allelic diversity. These results could explain why Venter's group found marine microbial genomes difficult to assemble from shotgun sequence data.

However, Kimura predicted that selection has more opportunity to act on small changes in fitness as population size increases, and therefore very large, stable populations should be more highly perfected by selection³⁶. More specifically, Kimura coined the term 'effective population size' to refer to the minimal size reached by a population that is undergoing fluctuations. For these marine microbes, it may be that where large populations have not been through recent episodes of purifying selection, they are able to maintain very large reservoirs of neutral genetic variation. If this hypothesis is correct, then, at least within ecotypes of microbial plankton, one would expect to find a core set of genes conferring relatively conserved phenotype.

The SAR11 clade provided the earliest demonstration that the subclades of environmental gene clusters could be ecotypes²⁶. Probing rRNA revealed the presence of a surface (IA) and a deep sub-clade (II), but failed to identify the niche of a third sub-clade (IB; ref. 26). More recently, the niche of SAR11 subclade IB emerged in a study of the transition between spring-bloom and summer-stratified conditions in the western Sargasso Sea²⁷. Nonmetric multidimensional scaling revealed that the IB subclade occurs throughout the water column in the spring, apparently giving way to the more specialized IA and II subclades when the water column becomes thermally stratified.

The ecotype concept continues to expand with the recognition that many microbial groups can be subdivided according to their distributions in the water column. The unicellular cyanobacteria are by far the best example. Two ecotypes of *Prochlorococcus* can readily be differentiated by their chlorophyll *b*/chlorophyll *a* ratios — a high-light-adapted (high-*b/a*) lineage, and a low-light-adapted (low-*b/a*) lineage. Phylogenetic evidence from internal transcribed spacers (ITS) suggests that the high-*b/a* strains can be differentiated into four genetically distinct lineages. The ITS-based phylogenies indicate that Marine Cluster A *Synechococcus* can be subdivided into six clades, three of which can be associated with adaptively important phenotypic characteristics (motility, chromatic adaptation, and lack of phycoerythrin)^{37,38}. So far, the genome sequences available have provided ample support for the hypothesis that these ecotypes differ in characteristics that affect their ability to compete. Notably, the low-*b/a* strain SS120 has a much smaller genome than the others and can use only ammonium and amino acids for nitrogen sources³⁹. At the other extreme, *Synechococcus* WH8102 can use ammonium, urea, nitrite, nitrate, cyanate, peptides and amino acids as sources of nitrogen. It is interesting to note that Marine Cluster A *Synechococcus* populations seem to prosper during periods of upwelling and vertical mixing — whereby nutrients are supplied but also cause chaotic, transitional conditions. Thus, as observed in the SAR11 clade, there seem to be seasonal specialists and stratification specialists in the marine unicellular cyanobacteria.

The observation that the major microbial plankton clades have diverged into ecotypes is powerful evidence that selection is creating functionally and genetically unique entities, despite the confounding influence of neutral variation, which causes relatively marked divergence in genome sequences. Although the unicellular marine cyanobacteria are a good model for what the future may hold, the debate about diversity is far from over.

Old paradigms challenged by new forms of phototrophy

The new millennium arrived in tandem with discoveries of new forms of phototrophy in the ocean surface, which in turn fundamentally changed perspectives on microbial food webs. Béja *et al.* reported the

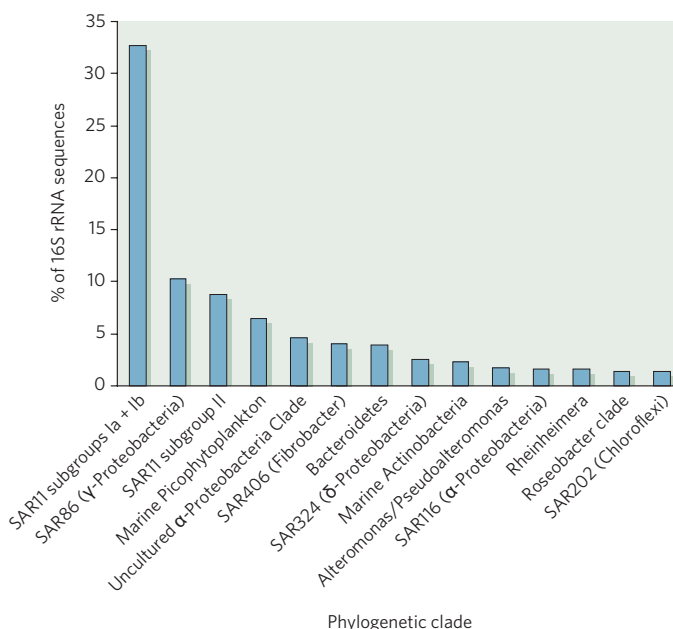


Figure 2 | 16S rRNA genes from the Sargasso Sea metagenome data set, organized by clades. The clades are shown in rank order according to gene abundance. To create the figure, 16S gene fragments were recovered from the data set by the BLAST program using 14 full sequences of different prokaryotic phyla as query. The resulting 934 sequences were phylogenetically analysed using the program package ARB. Venter *et al.* correctly reported more 16S rRNA gene fragments (1,184) because their analysis included smaller fragments that are excluded from the set of 934 sequences used in the analysis shown here. Genes belonging to the genera *Burkholderia* and *Shewanella* were omitted from the analysis because of suggestions that they are contaminants⁷².

presence of bacteriorhodopsin gene homologues that were linked to SAR86 rRNA genes on bacterial artificial chromosome (BAC) clones from Monterey Bay⁴⁰. And Kölber *et al.* used fast repetition rate infrared fluorescence spectroscopy to identify transient fluorescence signatures emanating from bacteriochlorophyll *a* in seawater⁴¹. Neither of these photosynthetic systems was biochemically novel; both had been previously discovered in obscure microbial groups. The two analogous systems use energy from light to pump protons out of cells, and are believed to function by charging transmembrane potentials and supporting uphill reactions, such as active transport and adenine dinucleotide phosphate (ADP) phosphorylation. What was new was the idea that bacterioplankton might be using these strategies on a vast scale to channel energy into cells that had been ostensibly regarded as respiratory heterotrophs. The details of which species have these systems, when they are expressed and what functions they support are obscure.

The very high abundance and diversity of bacteriorhodopsin homologues in the surface waters of the Sargasso Sea suggests that these genes are important¹⁸. Venter *et al.* recovered 743 'proteorhodopsin-like' sequences and 1,164 rRNA sequences from the Sargasso Sea — a ratio of 0.7. But, so far, only one report provides direct evidence that proteorhodopsins are active in seawater⁴²; another shows that incubating seawater in the dark has little impact on microbial populations, although this does inhibit the growth of the unicellular marine cyanobacteria^{43,44}. A key question is whether the proteorhodopsin gene family serves other functions as well — homologues of proteorhodopsin are also known to function as sensory receptors and to pump ions other than protons.

Kölber's report of the widespread presence of photosynthetic electron-transport systems based on bacteriochlorophyll *a* in the oceans was validated, but recent reports indicate that the quantitative significance of this process is less than was originally indicated⁴³.

Adaptive strategies

Some evidence suggests that bacteria can be divided between two general ecological strategies: those that exploit patchiness in the environment, and those that compete effectively at ambient nutrient background levels^{45,46}. Most of the strains studied by pioneering marine microbiologists were heterotrophic γ -Proteobacteria that grew very rapidly on marine nutrient agar. These organisms are rarely observed in 16S rRNA gene diversity surveys, but they frequently dominate experiments in which natural microbial plankton communities are manipulated^{47,48}.

Challenged with the problem of obtaining biologically relevant cultures, marine microbiologists have increasingly focused their efforts on new heterotrophic isolates that replicate effectively at very low nutrient concentrations. These organisms, referred to as oligotrophs, often do not respond well to high nutrient concentrations. One theory is that oligotrophs lack the complex regulatory responses and alternative metabolic pathways that enable their more robust counterparts to grow rapidly in high-nutrient media. The first isolate of SAR11, *Pelagibacter ubique*, grows slowly on seawater, doubling its population size every 29 hours, and does not increase its growth rate in response to the addition of organic carbon nutrients⁴⁹. Its genome is small — 1.3 million base pairs. There is also a suggestion that fast- and slow-growing bacteria may differ fundamentally in the rate of translation per ribosome⁵⁰.

Metabolic modelling has not 'arrived' in marine microbiology yet, but is looming on the horizon. Flux models for geochemical processes at the ecosystem level are historically one of the central themes of biological oceanography. It seems only a matter of time before the post-genomic effort to create *in silico* models for cells merges with the ecosystem flux models. In an intriguing foray into theory, Button *et al.* developed a mathematical model for oligotrophic cells that predicts their properties from basic principles⁵¹. The central concept in this theory is that evolution has acted on oligotrophic cell design to maxi-

mize a term called the specific affinity constant. The specific affinity constant takes into account the cell's overall ability to transform nutrients into biomass by considering the number of transporters and the surface area of the cell, in addition to transporter substrate affinities. This theory predicts that selection acts to reduce the size of oligotrophs so that the capacity of the cytoplasm to process substrate will not exceed the capacity of transport systems to supply substrate.

The concepts outlined above are general and do not take into account specialized adaptations that allow organisms to exploit specific features of the ecosystem. For example, most marine bacterioplankton seem to be lone cells that drift in the water column, but there is a subset of marine microbial groups that specialize in the colonization of particles. These cells have a special geochemical significance because particles sink, carrying carbon and other nutrients from the surface into the mesopelagic. This is the essence of the 'biological pump', which effectively reduces the partial pressure of carbon dioxide in the ocean surface, thereby increasing the rate at which atmospheric CO₂ dissolves in seawater⁵². A particularly important class of particles are macroscopic aggregates known as marine snow, which tend to form in the lower photic zone and disappear as the particles sink into the upper mesopelagic. DeLong *et al.* investigated organisms associated with marine snow and found little overlap between these organisms and the species freely suspended in the water column⁵³. Members of the Planctobacteria and Bacteroidetes phyla are particularly abundant on marine snow particles. Most Planctobacteria have specialized holdfast structures for attachment to surfaces. Bacteroidetes members frequently use gliding motility to traverse surfaces, and have been implicated in degradation of biopolymers associated with detritus⁵⁴. The recently discovered phylum Lentisphaerae has an unusual phenotype that suggests it might participate in particle formation⁵⁵. Lentisphaerae form a three-dimensional polysaccharide net that makes water viscous. A function for this net has not been proven, but the hypothesis that it serves to trap sinking particles fits the location of this organism, which resides in the upper mesopelagic, just beneath the photic zone.

Viral mortality and predator-prey relationships dominate current thinking about interactions between microbial plankton species⁵⁶, but it is difficult to believe that the water column has not spawned more insidious biological devices for gaining on the competition. A number of recent reports have focused on interactions between bacterioplankton and phytoplankton⁵⁷. Several strains of algicidal bacteria that produce algicidal peptides, antifouling agents or antimicrobial peptides have been described^{58,59}. The common genus *Pseudoalteromonas* crops up repeatedly in this context. But, for now, the passive 'microbial loop' model, in which heterotrophic microbial plankton acquire organic carbon that is lost from phytoplankton by leakage or lysis, reflects the prevailing view.

Rare species can be important

Some rare organisms that do not often show up in gene clone libraries may contribute significantly to oceanic geochemical processes. For example, the rate of nitrogen fixation is low in the oceans because fixed nitrogen is recycled efficiently in the photic zone, and other limiting nutrients prevent the expansion of plankton communities. However, nitrogen fixation plays a critical role in supporting productivity by replacing fixed nitrogen that is lost from the ocean surface. Zehr and colleagues identified diverse nitrogenase genes in seawater by culture-independent approaches^{60,61}. On the basis of Zehr's evidence, Montoya *et al.* identified new unicellular cyanobacteria that seem to contribute significantly to nitrogen fixation in the oceans⁶².

Cultivating key species

New microorganisms in culture are big news if they represent microbial groups that play important roles in the environment. Heightened attention to cultivation is producing a steady flow of new isolates from several research groups. The value of cultures has been underscored by

the problems that Venter's group encountered when reconstructing genomes from the Sargasso Sea metagenome data set. Complete genomes from cultures are now perceived as crucial reference points for assembling metagenome data. More importantly, strains in culture provide the means to study cell physiology and test hypotheses emerging from genome sequence data. Once again, the unicellular marine oxygenic phototrophs provide the best example. The cultivation of marine *Synechococcus* and *Prochlorococcus* species led to rapid advances in understanding the biology of these organisms (see above). Another group that has been studied in culture is the *Roseobacter* clade. One *Roseobacter* clade member, *Sulfitobacter*, has been shown to use dimethylsulphoniopropionate and can produce metabolic energy by sulphide oxidation⁶³. Another cyanobacterium that defied cultivation for decades, *Trichodesmium*, is now in culture in labs around the world⁶⁴.

Isolating cells by dilution (extinction culturing) into natural seawater was pioneered by Button *et al.*⁶⁵ and was later used to obtain *Sphingopyxis alaskensis*⁶⁶ and *Marinobacter arcticus*, which became model organisms for studying the properties of oligotrophic cells^{51,67}. High-throughput versions of extinction culturing led to the isolation of many more oligotrophic species^{55,68–70}, most of which replicate poorly or not at all on marine nutrient agar. Some of the important isolates to emerge from this approach are *Pelagibacter ubique*⁴⁹, the first cultured member of the SAR11 clade, as well as OM43, *Lentisphaera*⁵⁵ and the Oligotrophic Marine γ -Proteobacteria group⁶⁸. However, many of the dominant marine plankton groups, including the SAR86 cluster, the SAR202 cluster, the SAR324 cluster, the marine Actinobacteria, SAR406, and the group II marine Archaea, have not yet been cultivated. Given the number of replications of extinction cultures from seawater that have now been examined, it is clear that some of the important clades that remain uncultivated are not going to grow without further advances in cultivation technology.

Future prospects

Historically, biological oceanography has focused on developing quantitative models of food-web dynamics. With the emergence of the 'microbial loop' hypothesis in 1983 (ref. 71), bacteria assumed a major role in food-web models, but were treated as a 'black box' for the purpose of creating tractable models for geochemical fluxes. At a conceptual level the microbial black box has been repeatedly transformed by discoveries of new photosynthetic microorganisms and new forms of photoheterotrophy. If the traditions of oceanography eventually merge with current trends in microbiology, then the future metabolic models for the major bacterioplankton clades may become elements in larger models that predict geochemical fluxes in the ocean water column.

1. Michaels, A. F. *et al.* Seasonal patterns of ocean biochemistry at the U.S. JGOFS Bermuda Atlantic Time Series study site. *Deep-Sea Res.* **41**, 1013–1038 (1994).
2. Ducklow, H. in *Microbial ecology of the oceans* (ed. Kirchman, D. L.) 85–120 (Wiley-Liss, New York, 2000).
3. Fuhrman, J. A. & Noble, R. T. Viruses and protists cause similar bacterial mortality in coastal seawater. *Limnol. Oceanogr.* **40**, 1236–1242 (1995).
4. Carlson, C. A., Ducklow, H. W. & Michaels, A. F. Annual flux of dissolved organic carbon from the euphotic zone in the northwestern Sargasso Sea. *Nature* **371**, 405–408 (1994).
5. Bauer, J. E., Williams, P. M. & Druffel, E. R. M. ¹⁴C activity of dissolved organic carbon fractions in the north-central Pacific and Sargasso Sea. *Nature* **357**, 667–670 (1992).
7. Benner, R. in *Biogeochemistry of Marine Dissolved Organic Matter* (eds Hansell, D. A. & Carlson, C. A.) 59–85 (Academic, 2002).
6. Hansell, D. A. & Carlson, C. A. Biogeochemistry of total organic carbon and nitrogen in the Sargasso Sea: control by convective overturn. *Deep-Sea Res. Part II-Topical Studies in Oceanogr.* **48**, 1649–1667 (2001).
8. Baumann, P. & Schubert, R. H. W. in *Bergey's Manual of Systematic Bacteriology* (eds Krieg, N. R. & Holt, J. G.) 518–544 (Williams & Wilkins, Baltimore, 1984).
9. Giovannoni, S. J. & Rappé, M. S. in *Microbial Ecology of the Oceans* (ed. Kirchman, D. L.) 47–85 (Wiley-Liss, New York, 2000).
10. Britschgi, T. B. & Giovannoni, S. J. Phylogenetic analysis of a natural marine bacterioplankton population by rRNA gene cloning and sequencing. *Appl. Environ. Microbiol.* **57**, 1313–1318 (1991).
11. Fuhrman, J. A., McCallum, K. & Davis, A. A. Novel major archaeobacterial group from marine plankton. *Nature* **356**, 148–149 (1992).

12. Fuhrman, J. A., McCallum, K. & Davis, A. A. Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific oceans. *Appl. Environ. Microbiol.* **59**, 1294–1302 (1993).
13. Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. & Field, K. G. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**, 60–63 (1990).
14. Schmidt, T. M., DeLong, E. F. & Pace, N. R. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* **173**, 4371–4378 (1991).
15. Mullins, T. D., Britschgi, T. B., Krest, R. L. & Giovannoni, S. J. Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. *Limnol. Oceanogr.* **40**, 148–158 (1995).
16. DeLong, E. F. & Karl, D. M. Perspectives in microbial oceanography. *Nature* **437**, 336–342.
17. Acinas, S. G. *et al.* Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430**, 551–554 (2004).
18. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
19. Eder, W., Schmidt, M., Koch, M., Garbe-Schonberg, D. & Huber, R. Prokaryotic phylogenetic diversity and corresponding geochemical data of the brine-seawater interface of the Shaban Deep, Red Sea. *Environ. Microbiol.* **4**, 758–763 (2002).
20. Fuhrman, J. A., McCallum, K. & Davis, A. A. Widespread archaea and novel bacteria from the deep sea as shown by 16S rRNA gene sequences. *Mar. Ecol. Prog. Ser.* **150**, 275–285 (1997).
21. Giovannoni, S., Rappé, M., Vergin, K. & Adair, N. 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the Green Non-Sulfur bacteria. *Proc. Natl Acad. Sci. USA* **93**, 7979–7984 (1996).
22. Gordon, D. A. & Giovannoni, S. J. Stratified microbial populations related to *Chlorobium* and *Fibrobacter* detected in the Atlantic and Pacific Oceans. *Appl. Environ. Microbiol.* **62**, 1171–1177 (1996).
23. Wright, T. D., Vergin, K. L., Boyd, P. W. & Giovannoni, S. J. A novel delta-subdivision proteobacterial lineage from the lower ocean surface layer. *Appl. Environ. Microbiol.* **63**, 1441–1448 (1997).
24. Karner, M. B., DeLong, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507–510 (2001).
25. Morris, R. M. *et al.* SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**, 806–810 (2002).
26. Field, K. G. *et al.* Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl. Environ. Microbiol.* **61**, 63–70 (1997).
27. Morris, R. M., Cho, J. C., Rappé, M. S., Vergin, K. L. & Carlson, C. A. Bacterioplankton responses to deep seasonal mixing in the Sargasso Sea. *Limnol. Oceanogr.* **50**, 382–391 (2005).
28. Rappé, M. S., Vergin, K. & Giovannoni, S. J. Phylogenetic comparisons of a coastal bacterioplankton community with its counterparts in open ocean and freshwater systems. *FEMS Microbiol. Ecol.* **33**, 219–232 (2000).
29. Wuchter, C., Schouten, S., Boschker, H. T. & Sinninghe Damste, J. S. Bicarbonate uptake by marine Crenarchaeota. *FEMS Microbiol. Lett.* **219**, 203–207 (2003).
30. DeLong, E. F. Archaea in coastal marine bacterioplankton. *Proc. Natl Acad. Sci. USA* **89**, 5685–5689 (1992).
31. Zubkov, M. V. *et al.* Linking the composition of bacterioplankton to rapid turnover of dissolved dimethylsulphoniopropionate in an algal bloom in the North Sea. *Environ. Microbiol.* **3**, 304–311 (2001).
32. Zubkov, M. V., Fuchs, B. M., Tarran, G. A., Burkil, P. H. & Amann, R. High rate of uptake of organic nitrogen compounds by *Prochlorococcus* cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Appl. Environ. Microbiol.* **69**, 1299–1304 (2003).
33. Malmstrom, R. R., Kiene, R. P., Cottrell, M. T. & Kirchman, D. L. Contribution of SAR11 bacteria to dissolved dimethylsulphoniopropionate and amino acid uptake in the North Atlantic ocean. *Appl. Environ. Microbiol.* **70**, 4129–4135 (2004).
34. Cohan, F. M. What are bacterial species? *Annu. Rev. Microbiol.* **56**, 457–487 (2002).
35. Thompson, J. R. *et al.* Diversity and dynamics of a north Atlantic coastal *Vibrio* community. *Appl. Environ. Microbiol.* **70**, 4103–4110 (2004).
36. Kimura, M. On the probability of fixation of mutant genes in populations. *Genetics* **47**, 713–719 (1962).
37. Moore, L. R., Rocap, G. & Chisholm, S. W. Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**, 464–467 (1998).
38. Rocap, G., Distel, D. L., Waterbury, J. B. & Chisholm, S. W. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S–23S ribosomal DNA internal transcribed spacer sequences. *Appl. Environ. Microbiol.* **68**, 1180–1191 (2002).
39. Rocap, G. *et al.* Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**, 1042–1047 (2003).
40. Bèjà, O. *et al.* Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906 (2000).
41. Kölber, Z. S., Van Dover, C. L., Niederman, R. A. & Falkowski, P. G. Bacterial photosynthesis in surface waters of the open ocean. *Nature* **407**, 177–179 (2000).
42. Bèjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M. & DeLong, E. F. Proteorhodopsin phototrophy in the ocean. *Nature* **411**, 786–789 (2001).
43. Schwalbach, M. S., Brown, M. & Fuhrman, J. A. Impact of light on marine bacterioplankton community structure. *Aquat. Microb. Ecol.* **39**, 235–245 (2005).
44. Man, D. *et al.* Diversification and spectral tuning in marine proteorhodopsins. *EMBO J.* **22**, 1725–1731 (2003).
45. Stevenson, B. S. & Schmidt, T. M. Life history implications of rRNA gene copy number in *Escherichia coli*. *Appl. Environ. Microbiol.* **70**, 6670–6677 (2004).
46. Azam, F. & Long, R. A. Sea snow microcosms. *Nature* **414**, 495–498 (2001).
47. Carlson, C. A. *et al.* Effect of nutrient amendments on bacterioplankton production, community structure, and DOC utilization in the northwestern Sargasso Sea. *Aquat. Microb. Ecol.* **30**, 19–36 (2002).
48. Fuchs, B. M., Zubkov, M. V., Sahm, K., Burkil, P. H. & Amann, R. Changes in community composition during dilution cultures of marine bacterioplankton as assessed by flow cytometric and molecular biological techniques. *Environ. Microbiol.* **2**, 191–201 (2000).

49. Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**, 630–633 (2002).
50. Dethlefsen, L. & Schmidt, T. M. Differences in codon bias cannot explain differences in translational power among microbes. *BMC Bioinformatics* **6**, 3 (2005).
51. Button, D. K., Robertson, B., Gustafson, E. & Zhao, X. Experimental and theoretical bases of specific affinity, a cytoarchitecture-based formulation of nutrient collection proposed to supercede the Michaelis–Menten paradigm of microbial kinetics. *Appl. Environ. Microbiol.* **70**, 5511–5521 (2004).
52. Kohfeld, K. E., Le Quere, C., Harrison, S. P. & Anderson, R. F. Role of marine biology in glacial-interglacial CO₂ cycles. *Science* **308**, 74–78 (2005).
53. DeLong, E. F., Franks, D. G. & Alldredge, A. L. Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnol. Oceanogr.* **38**, 924–934 (1993).
54. Kirchman, D. L. The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS Microbiol. Ecol.* **39**, 91–100 (2002).
55. Cho, J. C., Vergin, K. L., Morris, R. M. & Giovannoni, S. J. *Lentisphaera araneosa* gen. nov., sp. nov., a transparent exopolymer producing marine bacterium, and the description of a novel bacterial phylum, *Lentisphaerae*. *Environ. Microbiol.* **6**, 611–621 (2004).
56. Suttle, C. Viruses in the sea. *Nature* **437**, 356–361 (2005).
57. Mayali, X. & Azam, F. Algicidal bacteria in the sea and their impact on algal blooms. *J. Eukaryot. Microbiol.* **51**, 139–144 (2004).
58. Egan, S., Thomas, T., Holmstrom, C. & Kjelleberg, S. Phylogenetic relationship and antifouling activity of bacterial epiphytes from the marine alga *Ulva lactuca*. *Environ. Microbiol.* **2**, 343–347 (2000).
59. Lovejoy, C., Bowman, J. P. & Hallingraeff, G. M. Algicidal effects of a novel marine *Pseudoalteromonas* isolate (Class Proteobacteria, Gamma subdivision) on harmful algal bloom species of the genera *Chattonella*, *Gymnodinium* and *Heterosigma*. *Appl. Environ. Microbiol.* **64**, 2806–2813 (1998).
60. Jenkins, B. D., Steward, G. F., Short, S. M., Ward, B. B. & Zehr, J. P. Fingerprinting diazotroph communities in the Chesapeake Bay by using a DNA microarray. *Appl. Environ. Microbiol.* **70**, 1767–1776 (2004).
61. Zehr, J. P., Jenkins, B. D., Short, S. M. & Steward, G. F. Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environ. Microbiol.* **5**, 539–554 (2003).
62. Montoya, J. P. et al. High rates of N₂ fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature* **430**, 1027–1032 (2004).
63. Moran, M. A. et al. Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* **432**, 910–913 (2004).
64. Orcutt, K. M. et al. Characterization of *Trichodesmium* spp. by genetic techniques. *Appl. Environ. Microbiol.* **68**, 2236–2245 (2002).
65. Button, D. K., Schut, F., Quang, P., Martin, R. & Robertson, B. R. Viability and isolation of marine bacteria by dilution culture: theory, procedures, and initial results. *Appl. Environ. Microbiol.* **59**, 881–891 (1993).
66. Vancanneyt, M. et al. *Sphingomonas alaskensis* sp. nov., a dominant bacterium from a marine oligotrophic environment. *Int. J. Syst. Evol. Microbiol.* **51**, 73–79 (2001).
67. Cavicchioli, R., Ostrowski, M., Fegatella, F., Goodchild, A. & Guixa-Boixereu, N. Life under nutrient limitation in oligotrophic marine environments: an eco/physiological perspective of *Sphingopyxis alaskensis* (formerly *Sphingomonas alaskensis*). *Microb. Ecol.* **45**, 203–217 (2003).
68. Cho, J. C. & Giovannoni, S. J. Cultivation and growth characteristics of a diverse group of oligotrophic marine γ -Proteobacteria. *Appl. Environ. Microbiol.* **70**, 432–40 (2004).
69. Connon, S. A. & Giovannoni, S. J. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl. Environ. Microbiol.* **68**, 3878–3885 (2002).
70. Simu, K. & Hagstrom, A. Oligotrophic bacterioplankton with a novel single-cell life strategy. *Appl. Environ. Microbiol.* **70**, 2445–2451 (2004).
71. Azam, F. et al. The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* **10**, 257–263 (1983).
72. DeLong, E. F. Microbial community genomics in the ocean. *Nature Rev. Microbiol.* **3**, 459–469 (2005).

Acknowledgements This work was supported by a research grant from the NSF Microbial Observatories Program and a grant from the Gordon and Betty Moore Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to S.J.G. (steve.giovannoni@oregonstate.edu).

Marine microorganisms and global nutrient cycles

Kevin R. Arrigo¹

The way that nutrients cycle through atmospheric, terrestrial, oceanic and associated biotic reservoirs can constrain rates of biological production and help structure ecosystems on land and in the sea. On a global scale, cycling of nutrients also affects the concentration of atmospheric carbon dioxide. Because of their capacity for rapid growth, marine microorganisms are a major component of global nutrient cycles.

Understanding what controls their distributions and their diverse suite of nutrient transformations is a major challenge facing contemporary biological oceanographers. What is emerging is an appreciation of the previously unknown degree of complexity within the marine microbial community.

To understand how carbon and nutrients, such as nitrogen and phosphorus, cycle through the atmosphere, land and oceans, we need a clearer picture of the underlying processes. This is particularly important in the face of increasing anthropogenic nutrient release and climate change. Marine microbes, which are responsible for approximately half of the Earth's primary production, play an enormous role in global nutrient cycling.

In this review, I will highlight the four exciting aspects of marine microbial ecology that are receiving a great deal of attention and may prove to be crucial to a revised understanding of marine and global nutrient cycling. The first involves the explanations for and consequences of the variable nutrient stoichiometry of phytoplankton. The second is the emerging concept that phytoplankton growth can be limited by more than one resource. The third concerns the upward revision of estimates of marine nitrogen fixation, and the fourth is the discovery that fixed nitrogen in the ocean can be lost through anaerobic ammonium oxidation (anammox) reactions. Although distinct, these topics all represent examples of the marine microbial community modulating the coupling between the cycles of nutrients and carbon. Consequently, they all have the capacity to fundamentally alter our perceptions of global nutrient cycles and their response to environmental change.

Non-Redfield behaviour of phytoplankton

In the early part of the twentieth century, Alfred Redfield noticed that the elemental composition of plankton was strikingly similar to that of the major dissolved nutrients in the deep ocean¹. On the basis of these observations, Redfield proposed that the nitrate:phosphate ($\text{NO}_3\text{:PO}_4$) ratio of 16:1 in the sea was controlled by the requirements of phytoplankton, which subsequently release nitrogen (N) and phosphorus (P) to the environment at this ratio as they are broken down (remineralized). Redfield's initial observations have been confirmed numerous times, and the notion of a 'Redfield ratio' describing the stoichiometry of both phytoplankton and seawater remains a fundamental tenet shaping our understanding of marine ecology, biogeochemistry and even phytoplankton evolution. The Redfield ratio has been extended to include other elements, most notably carbon (C), and it links these three major biogeochemical cycles

through the activities of marine phytoplankton.

Unfortunately, a clear mechanism explaining the observed magnitude of the Redfield C:N:P ratio of 106:16:1 for either phytoplankton or the deep ocean has been elusive. It has long been recognized that conditions exist under which phytoplankton stoichiometry diverges from the canonical Redfield ratio. Furthermore, a number of processes drive oceanic nutrient inventories away from the Redfield ratio, including changes in exogenous nutrient delivery² and microbial metabolism³ (for example, nitrogen fixation, denitrification and anammox, see below). These processes are sometimes manifested as variations in N^* , a measure of the degree of N deficit or excess relative to P for a given water mass⁴. What governs variations in phytoplankton nutrient stoichiometry, and, given that variation, why is the Redfield N:P ratio observed in the deep ocean so universal?

At the most basic level, the C:N:P stoichiometry of extant phytoplankton reflects the elemental composition retained from their early evolutionary history⁵. In the case of eukaryotic phytoplankton, the two major superfamilies differ markedly in their cellular C:P and N:P ratios, with the green superfamily exhibiting significantly higher ratios than the red (green, C:P $\approx$ 200 and N:P $\approx$ 27; red, C:P $\approx$ 70 and N:P $\approx$ 10). However, all observed C:N:P stoichiometries cannot be explained by the evolutionary lineage of an organism. The highly dynamic stoichiometry often exhibited by unicellular algae reflects their ability to store nutrients in internal pools, switch between enzymes with different nutrient requirements and modify osmolyte composition^{6,7}. Lower-frequency variations in C:N:P stoichiometry are related to changes in the structural elements of the phytoplankton cell. A major breakthrough in our understanding of cellular C:N:P stoichiometry came with the realization that different cellular components have their own unique stoichiometric properties. Most notably, resource (light or nutrients) acquisition machinery, such as proteins and chlorophyll, is high in N but low in P, whereas growth machinery, such as ribosomal RNA, is high in both N and P^{8,9}. Because these components make up a large proportion of cellular material, changes in their relative proportions have a marked effect on bulk cellular C:N:P stoichiometry. Why then might the proportions of these components change?

¹Department of Geophysics, Stanford University, Stanford, California 94305-2215, USA.

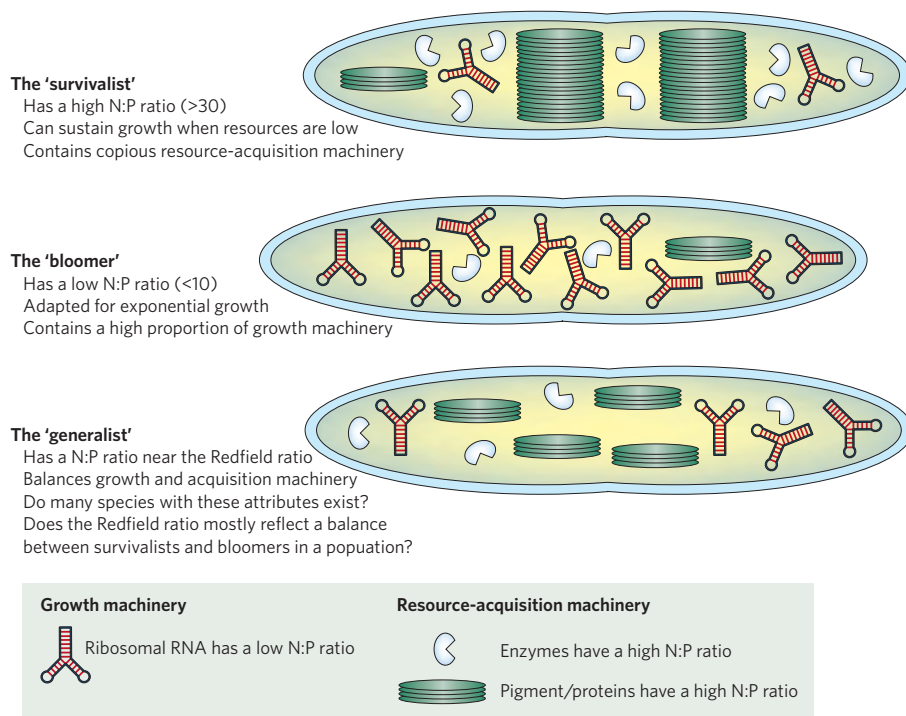


Figure 1 | Three different phytoplankton growth strategies and their resulting cellular N:P ratios. The allocation of resources and resulting N:P ratios for the 'survivalist' and the 'bloomer' are from the optimization model of Klausmeier *et al.*¹⁰. Interestingly, the model does not predict an optimal N:P ratio of 16 (our hypothetical 'generalist') under any of the environmental conditions tested. This indicates that the Redfield N:P ratio of 16 observed in nature is simply an average value that reflects an ecological balance between the 'survivalists' and 'bloomers' in a population.

This question was elegantly addressed recently using an optimization model¹⁰ based on relative shifts in the cellular machinery described above (Fig. 1). It predicts that during exponential growth, bloom-forming phytoplankton optimally increase their allocation of resources toward production of growth machinery, reducing their N:P ratio to ~8, far below the Redfield value of 16. However, when resources are scarce, slow-growing phytoplankton that can synthesize additional resource acquisition machinery are favoured. This allocation of resources results in optimal N:P ratios ranging from 36 to 45, depending on which resource is limiting. Results of this model agree remarkably well with observed variations in N:P ratios reported for dozens of phytoplankton taxa growing under different environmental conditions¹⁰. The model is even able to explain the unusually high N:P ratio that has been measured for dinitrogen (N_2) fixers (> 40). Rather than simply being the result of a ready-made N supply, it now appears that the high cellular N:P ratio of N_2 fixers is best explained by their need for large quantities of P-poor light-harvesting machinery required to drive N_2 fixation¹¹.

The key conclusion of the optimization model is that the canonical Redfield N:P ratio of 16 for phytoplankton is not a universally optimal value but instead represents an average for a diverse oceanic phytoplankton assemblage growing under a variety of different conditions and employing a range of growth strategies (Fig. 1). Consequently, the deep-sea $NO_3:PO_4$ ratio of ~16:1 simply reflects the stoichiometry of the current global phytoplankton community. This implies that, as environmental conditions change, the global mean phytoplankton nutrient stoichiometry could vary over time, potentially modifying current nutrient inventories. One change observed recently has been an increase in the flux of iron (Fe)-rich dust from an increasingly desertified Asian continent to the North Pacific ocean¹². The response of the microbial community to this Fe enrichment has been enhanced particulate C:P and C:N ratios, reflecting increased export production and associated rates of remineralization¹³. Similarly, the N:P ratio of sinking particulate matter in the North Atlantic ocean has increased over the past 50 years, possibly because of increased N availability through atmospheric deposition of anthropogenically produced nitrous oxides¹³.

In addition, if predictions by global climate models (GCMs) are correct and if surface waters in polar regions become more stratified¹⁴,

then the composition of phytoplankton species could be markedly altered¹⁵. Over time, a shift towards an increase in the abundance of diatoms, which prefer stratified waters and have a much lower N:P requirement¹⁵ than the other major polar phytoplankton taxon, *Phaeocystis*¹⁶, could deplete NO_3 relative to PO_4 and reduce the N:P ratio of subsurface waters, which are currently near the Redfield ratio¹⁷. NO_3 depletion may be further amplified by relatively high rates of P recycling in surface waters. Because waters formed in the Antarctic circulate globally¹⁸, such long-term shifts in phytoplankton species composition and nutrient uptake characteristics could have implications for nutrient inventories in waters around the world.

Gaining a better understanding of microbial C:N:P stoichiometry is essential because of the predominant role these largely biologically mediated relationships play in coupled elemental cycles. These cycles modulate, and are themselves modulated by, processes operating at scales ranging from algal photosynthesis to the global climate. Unfortunately, although the amount of atmospheric carbon dioxide removed by the ocean in GCMs is very sensitive to the stoichiometric relationships between phytoplankton and nutrients, so far, few models account for its observed variability. From a more human perspective, stoichiometric relationships also determine how the marine environment responds to increasing anthropogenic inputs of limiting nutrients, for example, the increase in nitrogenous fertilizers¹⁹. Finally, as will be discussed below, the balance between nutrient inventories and the stoichiometric requirements of cells controls fundamental aspects of marine microbial ecology and biogeochemistry, such as where and how much N_2 fixation, denitrification and anammox take place and what nutrients limit phytoplankton growth (Fig. 2).

Multiple resource co-limitation

For many decades, Liebig's law of the minimum²⁰, which states that only a single resource limits plant growth at any given time, was a dominant theory shaping how oceanographers viewed phytoplankton ecology and its impact on nutrient cycles. During the past decade, this simple view has been replaced by the realization that in some parts of the world's ocean multiple resources simultaneously limit phytoplankton growth. Resource co-limitation is a phenomenon observed most commonly in oligotrophic oceans (Fig. 2), where nutrients are low but are nevertheless responsible for a large

fraction of global net primary production²¹.

Although resource co-limitation has been recognized for only a short time, it has already been attributed to a variety of situations. Hence, its precise definition remains unclear. Resource co-limitation has been invoked when phytoplankton growth is stimulated either by the simultaneous addition of two or more different resources (both resource A and B are limiting) or by the addition of different individual resources (either resource A or B is limiting). Furthermore, resource co-limitation has been attributed to responses ranging from the cellular to the community level. Below I use examples gleaned from the literature to define three distinct categories of resource co-limitation that apply most often in the marine environment (Fig. 3).

The simplest case of resource co-limitation exists when two or more nutrients are reduced to levels too low for cellular uptake. This can happen, for example, when luxury uptake by phytoplankton preferentially depletes the more abundant nutrient or when N_2 fixers draw down PO_4 to growth-limiting levels when N is already limiting. In situations like these, referred to here as multi-nutrient co-limitation, adding all of the limiting nutrients is required for phytoplankton growth. Multi-nutrient co-limitation was observed in the nutrient-depleted waters of the Baltic Sea, where addition of both N and P was required to stimulate phytoplankton growth²². Similarly, Si and P were found to be co-limiting the growth of diatoms in the South China Sea near the Pearl River estuary²³. More recently, the simultaneous addition of both P and Fe was required to stimulate growth of N_2 -fixing cyanobacteria in the tropical North Atlantic²⁴.

A fundamentally different type of resource co-limitation arises when phytoplankton growth is stimulated by the addition of one of a number of different resources (the addition of either resource A or B increases phytoplankton growth). Here, although both resources may be limiting, only one is required to elicit a growth response. This can happen under one of two conditions (Fig. 3). In the first, referred to here as biochemical co-limitation, addition of one limiting resource may facilitate the uptake or assimilation of another (previously) limiting resource. In the second, called community co-limitation, one segment of the phytoplankton population may respond to an increase in one resource whereas another segment may respond to the increase in a different resource.

Figure 3 | A breakdown of the three types of resource co-limitation. In this simplest of examples, there are two resources (A and B) and two members of the phytoplankton community. Resource co-limitation has been invoked in cases when the addition of resources A and B are both required for phytoplankton growth (multi-nutrient co-limitation) and in cases when addition of either resource A or resource B stimulates phytoplankton growth. In the latter case, either the presence of resource A facilitates the assimilation of resource B (biochemical co-limitation) or one member of the phytoplankton community responds to resource A and the other member to resource B (community co-limitation).

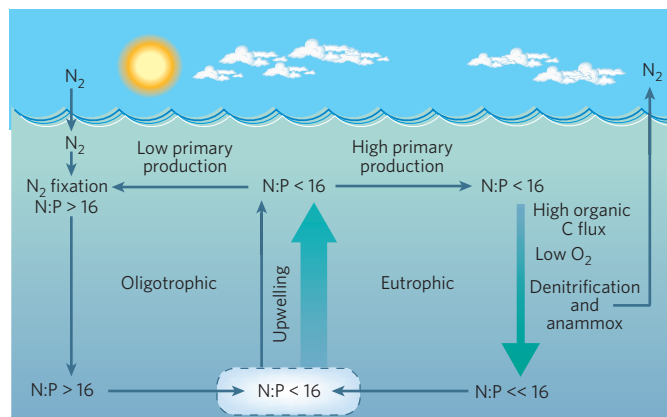
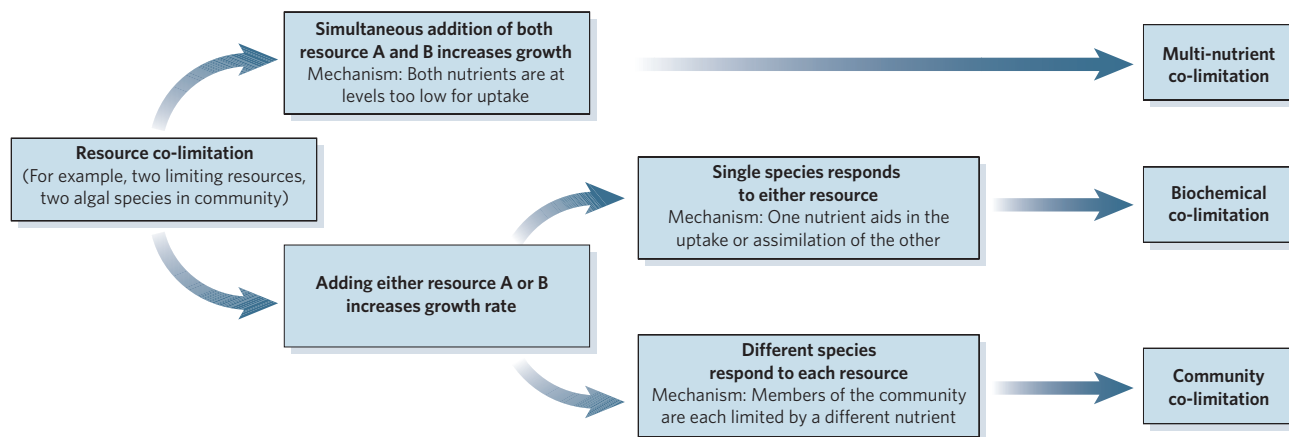


Figure 2 | The global ocean balance between N_2 fixation and the loss of fixed N through anammox and denitrification. Waters upwelling to the surface are generally slightly depleted in N relative to P (below Redfield). Where upwelling of N and P is substantial (eutrophic regions), primary production is high, resulting in the sinking of large amounts of organic matter. As this organic matter is broken down and its N and P are solubilized, a large fraction of the available O_2 is consumed. In these suboxic waters, anammox and denitrification converts NH_4 and NO_3 to N_2 , resulting in a loss of biologically reactive N from the system and a marked decrease in the deep ocean N:P ratio. In contrast, where upwelling of N and P is low (oligotrophic regions), primary production is reduced and often dominated by N_2 -fixing cyanobacteria, which are favoured in low-nutrient, N-depleted surface waters. N_2 fixation increases the N:P ratio of the organic matter to values above the Redfield ratio, which is eventually remineralized to produce waters enriched in N. However, global rates of N_2 fixation are not sufficient to balance the losses of fixed N to anammox and denitrification. Consequently, the contemporary ocean has a mean N:P ratio slightly less than the Redfield ratio.

In the case of biochemical co-limitation, the assimilation of resource A depends upon its abundance within the environment and the amount of cellular machinery available for its assimilation or transport into the cell. Resource B (often a trace metal cofactor) is integral to the functioning of the machinery required to assimilate or transport resource A. Hence, adding either resource increases phytoplankton growth. An important characteristic of biochemical co-limitation is that it operates entirely at the cellular level.

An early example of biochemical co-limitation involves the interaction between light and Fe. During culture²⁵ and field experiments²⁶, phytoplankton growth was enhanced either by increasing the light level or by adding Fe, an essential component of the additional photosynthetic units required for absorption of low light. Furthermore, the degree of biochemical co-limitation was shown to vary with cell size²⁷. Small-celled *Chaetoceros* species in the Southern Ocean were less sus-

ceptible to light limitation at low Fe concentrations owing to their high surface area to volume ratio and, thus, their superior ability to take up Fe. Larger *Chaetoceros* species could not acquire sufficient Fe to grow under similarly low light and only bloomed when Fe and light were abundant.

A more recent example of biochemical co-limitation²⁸ involves PO_4 and the trace element zinc (Zn). In waters with low inorganic P, some phytoplankton can access the more abundant dissolved organic P (DOP) pool²⁹. This is achieved through the activity of alkaline phosphatase (AP), an enzyme that cleaves the P-containing moiety from DOP so that it can be taken into the cell. However, AP requires a Zn cofactor. Consequently, the ability of phytoplankton to exploit DOP is compromised at reduced Zn concentrations. Experiments have shown that under such conditions, the rate of phytoplankton growth can be increased either through DOP enrichment, with the DOP being cleaved more rapidly at higher concentrations by AP, or by adding more Zn, which facilitates the production of more AP, thereby allowing the cells to more efficiently access the available DOP²⁸.

In the case of community co-limitation, adding one nutrient increases the growth of one segment of the population, whereas adding a different nutrient increases the growth of another. Thus, community co-limitation can occur even if individual members of the community are each limited by only a single nutrient. The best-known example of community co-limitation is found in Fe-replete oligotrophic waters where N_2 -fixing cyanobacteria are abundant and N and P are scarce. Addition of NO_3 to these waters stimulates non- N_2 fixers in the population capable of assimilating DOP using AP^{30} . Addition of P stimulates cyanobacterium such as *Trichodesmium* that support their high N requirement through N_2 fixation^{31,32}. Community co-limitation can even involve assemblages where some species are multi-nutrient co-limited whereas others are limited by a single nutrient. For instance, in the tropical north Atlantic, addition of both P and Fe is required to stimulate N_2 fixers²⁴. Addition of N alone stimulates non- N_2 -fixing picoplankton, such as *Prochlorococcus* and *Synechococcus*, which may be more effective than the N_2 fixers at using the low P and Fe concentrations found in these waters.

It is becoming increasingly clear that the old notion of a single nutrient (or other resource such as light) limiting the growth of marine phytoplankton must give way to a more complex view that allows for limitation by multiple nutrients, both at the level of the individual cell and at the level of the entire community. Because nutrient inventories in seawater are profoundly influenced by the activities within the microbial community on both short and long timescales, it is essential that we understand how interactions between multiple resources can influence community composition and the nature of nutrient cycling between particulate and dissolved phases. This will prove particularly important for understanding and predicting how a non-steady-state ocean^{13,31} responds to further anthropogenic influences that alter the nutrient stoichiometry of surface waters.

A new paradigm for nitrogen fixation

The balance between cellular nutrient stoichiometry and oceanic nutrient inventories often shapes phytoplankton community composition, with N_2 -fixing phytoplankton being favoured in well-lit tropical and subtropical waters depleted of inorganic N^{33} . N_2 fixation by these organisms fuels new production and helps determine the amount of C and N that can be consumed by other organisms or exported to depth. The importance of N_2 fixation for the marine N cycle has received considerable attention in the past decade, prompted in part by recent estimates of N demand in the euphotic zone that exceed NO_3 fluxes from depth³⁴. This upward revision in N demand should be balanced primarily by N_2 fixation, although increased atmospheric N deposition or non-Redfield stoichiometry may also play a role³⁵. However, early survey work suggested that marine N_2 fixers were relatively rare and N_2 fixation was of minor importance^{36,37}, amounting to only 10–20 Tg N yr^{-1} , far below the 90–130 Tg N yr^{-1}

fixed in the terrestrial environment³⁸. How then can this increased N-demand be satisfied if N_2 fixation is so low? The likely answer is that the abundance of the N_2 fixers such as *Trichodesmium*, the most common representative, has been severely underestimated³⁹ and that N_2 fixation is a more important component of the marine N cycle than previously realized.

Over timescales of hundreds to thousands of years^{40,41}, the amount of biological N_2 fixation approximately balances the slightly greater losses of fixed N due to microbial processes such as denitrification (Fig. 2). Although absolute rates of these two processes may be controlled through complex climate feedbacks involving changes in Fe availability⁴², the steady-state oceanic balance between N_2 fixation and denitrification is maintained primarily through the N:P ratio of phytoplankton (Fig. 2). Simply put, if rates of denitrification were to increase because of elevated Fe input, waters welling up to the surface would be depleted in N relative to P, favouring the growth of N_2 fixers until P became depleted⁴³. If rates of denitrification were to drop, N_2 fixation would cease once the N:P ratio of upwelled waters was again equal to the N:P requirements of the phytoplankton and N was no longer limiting.

N_2 fixation requires the Fe-rich nitrogenase enzyme complex^{11,41}, and so N_2 fixers were thought to have Fe requirements orders of magnitude greater than phytoplankton growing on ammonium (NH_4)⁴⁴. Consequently, N_2 fixation was considered to be possible only in surface waters receiving large aeolian Fe input^{32,45}. This notion was consistent with observed correlations between areas with high dust fluxes and *Trichodesmium* distributions³⁹. However, newer evidence, gathered under trace-metal-clean conditions^{46,47}, suggests that Fe requirements of N_2 fixers are actually only about two to five times greater than for phytoplankton growing on NH_4 and only slightly higher than phytoplankton growing on NO_3 . Accordingly, the notion that N_2 fixation is universally controlled by Fe availability is being replaced by a more complex view that allows for limitation by other nutrients such as PO_4 (ref. 46), or even co-limitation by P and Fe²⁴. This is particularly true in areas such as the oligotrophic Pacific³¹ and Atlantic⁴⁸ and the northern Red Sea⁴⁹, where aeolian dust deposition raises surface Fe concentrations to ~1 nM, levels generally non-limiting to N_2 fixation.

Although *Trichodesmium* is considered the dominant N_2 fixer in the ocean⁵⁰, other N_2 fixers also deserve consideration. Most notably, the diatom genera *Rhizosolenia* and *Hemiaulus* contain the endosymbiotic N_2 -fixing cyanobacteria *Richelia intracellularis*⁵¹, which is capable of extremely high N_2 -fixation rates. During one study in the western tropical north Atlantic, nearly 70% of N demand in surface waters was met by N_2 fixation during an extensive bloom of these organisms⁵². In addition, it has recently been discovered that some species of unicellular cyanobacteria and proteobacteria also fix N_2 (ref. 53). However, neither the distributions nor the N_2 -fixation rates of these organisms are well known⁵⁴, although molecular tools are being developed to better quantify the abundance of N_2 fixers in a phytoplankton population⁵⁵. In some regions, their contribution to total N_2 fixation is low and in others, such as the oligotrophic North Pacific, it can exceed that of *Trichodesmium*⁵⁶.

Extrapolating from daily N_2 -fixation rates to obtain global annual values has been difficult because of the large geographic area over which N_2 fixation is important and the relatively small number of direct measurements. Nevertheless, recent efforts are encouraging. Annual rates of N_2 fixation by *Trichodesmium* in the north Atlantic have been estimated to range between 22 and 34 Tg N yr^{-1} (ref. 34). These values are consistent with geochemical proxies suggesting that N_2 fixation adds an average of 28 Tg N yr^{-1} to the oceanic N inventory^{4,34}. Applying this type of analysis to the global ocean yields annual N_2 -fixation rates in the range of 100–200 Tg N yr^{-1} (refs 4,41). Although there are still large uncertainties in this value⁵⁷, it is clear that rates of oceanic N_2 fixation are much greater than the 10–20 Tg N yr^{-1} estimated previously^{36,37} and that they are at least as high as those measured for the terrestrial environment³⁸. In the subtropical and

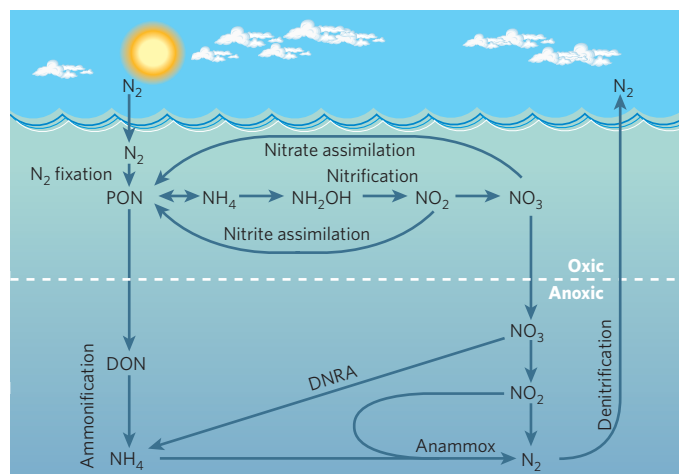


Figure 4 | Marine N cycle, including losses of ammonium and nitrite as N_2 owing to anammox. PON, particulate organic nitrogen, including phytoplankton; DON, dissolved organic nitrogen, DNRA, dissimilatory nitrate reductase to ammonium.

tropical Atlantic and Pacific oceans, N_2 fixation accounts for approximately 36–50% of the total N demand by the microbial community^{52,58}. It has been estimated that N_2 fixation is equivalent to 50–180% of the flux of NO_3 into the euphotic zone^{31,34}, demonstrating that a large fraction of the new production in these waters is actually fuelled by N_2 fixation, rather than by movement of NO_3 into the euphotic zone.

As our understanding of the global importance of N_2 fixation grows, other aspects of this unique process are coming to light. For example, oceanic rates of N_2 fixation may not be at steady state and can vary in intensity with changes in climate^{31,41}. These changes may include increased upper-ocean stratification, which could enhance N_2 fixation and shift organic matter export from being mostly particulate to mostly dissolved³¹. Consequently, changes in N_2 fixation can markedly alter the N inventory of the ocean and, hence, the stoichiometric balance between C, N and P available as nutrients. Yet despite its obvious importance, few large-scale biogeochemical models that include the tropics and subtropics include the process of N_2 fixation. More work in this area is necessary if we are to fully understand the future biogeochemical role of biologically fixed N_2 under the expected conditions of enhanced upper-ocean warming and stratification and, possibly, changes in dust fluxes.

Microbial N metabolism in low-oxygen environments

Denitrification is a microbial process thought to be responsible for the global loss of 175–450 Tg fixed N yr^{-1} within oxygen (O_2)-depleted environments^{4,59}. It is a crucial component of the marine N (Fig. 4) and C cycles because it reduces the N content of waters upwelling to the surface (Fig. 2), favouring the growth of N_2 fixers in oligotrophic waters. Years ago, Richards⁶⁰ noticed that most of the NH_4 that should be produced during the anaerobic remineralization of organic matter was unaccounted for. He proposed that the missing NH_4 was anaerobically oxidized to N_2 by some unknown microbe using NO_3 as an oxidant. Because there was no known biological pathway for this transformation, biological anaerobic NH_4 oxidation received little further attention.

By the mid-1990s, work with bioreactors designed to remove NH_4 from wastewater provided direct evidence for anaerobic ammonium oxidation, and the process was termed ‘anammox’⁶¹. In this reaction, NH_4 is oxidized to N_2 using nitrite (NO_2) as an oxidant⁶², consistent with Richards’ suggestion 30 years earlier. During anammox (Fig. 4), N_2 is formed through pairing of one N atom from both NO_2 and NH_4 (ref. 63), clearly distinguishing anammox from denitrification, which combines N from two NO_3 molecules to form N_2 . A few years after the discovery of anammox, the microbes responsible for the reaction were identified as

members of the bacterial order Planctomycetales⁶⁴. This discovery reinvigorated efforts to determine whether anammox is an important process controlling N distributions in the marine environment.

Anammox activity in the ocean was first investigated in anoxic sediments using ^{15}N -labelled NH_4 . The approach is based on the fact that because anammox combines N from NO_2 with NH_4 to form N_2 , the anammox reaction can be recognized by the production of singly labelled N_2 gas ($^{15}N^{14}N$, with the ^{14}N coming from NO_2 and the ^{15}N coming from NH_4). Surprisingly, this study⁶³ found that anammox accounted for 24–67% of the total N_2 production in the continental shelf sediments that were studied. (Denitrification accounted for the rest.)

Soon after, some of the first evidence of the anammox reaction in suboxic waters (rather than sediments) came from the Golfo Duce, Costa Rica⁶⁵. Using the ^{15}N -labelling technique, researchers showed that the observed NH_4 deficiency in low- O_2 waters was due to coupling between denitrification and anammox, with denitrification providing the NO_2 required by the anammox reaction. During this study, anammox accounted for 19–35% of total N_2 production, and as much as 58% at some depths. Anammox in oxygen minimum zones (OMZ) was estimated to account for 10–15% of the loss of fixed N from the world’s oceans. It was suggested that anammox may be even more important in regions such as the Peruvian and Chilean upwelling zones, where anoxic NO_3 -rich waters contact sediments that produce significant amounts of NH_4 .

Although anammox had been shown to be a potentially important process for removal of fixed N in O_2 -deficient marine waters, it had yet to be determined whether the organisms responsible were the same planctomycetes identified in sediments. This breakthrough was first achieved during a study of the suboxic and anoxic zones of the Black Sea³. Again using the ^{15}N -labelling technique, direct evidence of anammox activity was found below the oxic zone. Moreover, ladderane lipids were found in association with the zone of high anammox activity. These marker lipids are unique to the bacterial anammoxosome, the cellular structure where anammox reactions take place⁶⁶. This observation provided strong evidence that anammox bacteria, similar to those found living in bioreactors, were responsible for the anaerobic NH_4 oxidation observed in the Black Sea. Additional evidence was provided by the use of molecular probes that are specific for planctomycetes⁶⁷. These probes detected anammox bacteria in suboxic waters of both the central basin and the shelf break. The activity of these anammox bacteria was sufficient to oxidize all the NH_4 diffusing up into the suboxic zone and to consume > 40% of the fixed N sinking into anoxic waters of the Black Sea.

More recent studies have reported anammox activity in marine ecosystems as diverse as polar sea ice⁶⁸ and the OMZ of the Benguela upwelling system⁶⁹. Bottom waters in the Benguelan OMZ become suboxic because of O_2 decomposition of sinking organic matter after the highly productive upwelling season⁷⁰. The low levels of fixed N relative to PO_4 in these suboxic waters have generally been attributed to denitrification⁷¹. However, the low NO_3 and NH_4 in waters with reduced O_2 suggest that anammox may play a role. This role was recently confirmed by ^{15}N -labelling, which detected significant anammox activity⁶⁹. Additional analyses demonstrating the presence of both ladderane lipids and the 16S ribosomal RNA of a form of planctomycetes that is closely related to known anammox bacteria, verify that the loss of fixed N was due to anammox bacteria⁶⁹.

Among the most startling findings of the Benguela ONZ study came from the ^{15}N labelling experiments that showed that very little N_2 was produced by denitrifiers. The role of the denitrifiers in these waters seemed to be limited to reducing NO_3 to NO_2 , with anammox bacteria completing the process by converting NO_2 to N_2 during the oxidation of NH_4 . The authors of that study⁶⁹ boldly point out that they are aware of no direct evidence from ^{15}N labelling experiments that denitrifiers in OMZs anywhere in the world can convert NO_3 to N_2 . One possible reason is that nobody has bothered to look, given that,

until very recently, denitrification was the only known pathway capable of converting NO_3 to N_2 . However, if their suggestion is true, then anammox, perhaps coupled to denitrification, may be the dominant process in the loss of 80–150 Tg N yr^{-1} from the world's OMZs^{4,59}.

Studies of anammox are still in their infancy. What is becoming clear, however, is that this process is turning up in an ever-increasing number of low- O_2 marine environments. If it is prevalent in the Black Sea and the Benguela upwelling system, might it also be important in other areas favourable for denitrification, such as the west coasts of India and Central and South America⁷²? What about the recently discovered 'dead zones' near the coast of the northwest United States⁷³ or the anoxic waters of the Gulf of Mexico? Current estimates suggest that globally anammox may be responsible for 30–50% of N_2 production in the ocean⁷² and, if Kuypers *et al.*⁶⁹ are correct, much more. One important issue that needs to be addressed is the extent to which anammox represents an as-yet unquantified loss of fixed N. Will anammox alter existing N budgets because it consumes NH_4 rather than NO_3 , or is it already accounted for in estimates of denitrification? It will be interesting to watch as the anammox story continues to unfold.

Future directions

The past decade has seen profound shifts in some of the fundamental theories of biological oceanography, particularly as they relate to microbe–nutrient interactions. The simple concepts of a uniform Redfield stoichiometry for phytoplankton, single resources universally limiting phytoplankton growth, low levels of marine N_2 fixation and the pre-eminence of denitrification in the loss of fixed N are all being replaced by more complex conceptual models. It is also becoming increasingly clear that these seemingly disparate biological/biogeochemical concepts are really linked through the activity and elemental composition of marine microbes (Fig. 2). The C:N:P stoichiometry of phytoplankton ultimately controls the nutrient ratios of the deep ocean, which are subsequently modified by microbial reactions such as anammox and denitrification. The abundance and elemental ratio of nutrients returning to the surface determine the activity and composition of the phytoplankton, favouring rapid-growing taxa where nutrients are in abundance and N_2 fixers in nutrient-poor waters with a low N:P ratio (Fig. 4). These interactions produce a self-regulating biogeochemical system that maintains quasi-stable oceanic nutrient inventories over both short and long timescales.

Although our understanding of the roles of microbes in global nutrient cycles has increased markedly, the recent discovery of a whole new N biochemistry through anammox aptly illustrates that there is still much to be done. For instance, global biogeochemistry models that currently assume constant Redfield ratios may have to be reformulated to take into account a time-varying, taxon-specific non-Redfield stoichiometry. N budgets will need to be adjusted as we re-evaluate the balance between N_2 fixation and fixed N losses, as well as the relative magnitudes of denitrification and anammox.

Future work will almost certainly be driven by the development of improved tools for more comprehensive observation and quantification of microbial processes. Among other things, these new tools will facilitate an expansion of the *in situ* genomic work that is currently underway (see the review in this issue by DeLong and Karl, page 336) and that has allowed us to identify microbial genes governing a wide array of functions within a large number of marine habitats. For example, the widespread incidence of the *nrfA* genes that code for NO_2 reductase suggest that this pathway may be another important, yet largely unaccounted for, component of the N cycle⁷⁴. New tools also will improve characterization of marine microbial communities using satellite remote sensing technology. Currently, we use satellites to estimate phytoplankton biomass and productivity, and in some cases, distributions of taxa such as *Trichodesmium* and coccolithophores. As both sensor performance and our bio-optical characterization of different phytoplankton taxa improves, so will our ability to monitor changes in phytoplankton biomass and community structure. Finally,

new modelling tools will be able to incorporate our increasing understanding of the role of marine microbes in global nutrient cycling. Because many biological and biogeochemical processes involve complex interactions and multiple feedbacks, predicting their response to environmental perturbations is difficult, but essential, if we are to accurately characterize our ever-changing world. ■

1. Redfield, A. C. in *James Johnstone Memorial Volume* (ed. Daniel, R. J.) 176–192 (Liverpool Univ. Press, Liverpool, 1934).
2. Fanning, K. A. Influence of atmospheric pollution on nutrient limitation in the ocean. *Nature* **339**, 460–463 (1989).
3. Kuypers, M. M. M. *et al.* Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. *Nature* **422**, 608–611 (2003).
4. Gruber, N. & Sarmiento, J. L. Global patterns of marine nitrogen fixation and denitrification. *Global Biogeochem. Cycles* **11**, 235–266 (1997).
5. Quigg, A. *et al.* The evolutionary inheritance of elemental stoichiometry in marine phytoplankton. *Nature* **425**, 291–294 (2003).
6. Sterner, R. W. & Elser, J. J. *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere* (Princeton Univ. Press, Princeton, 2002).
7. Klausmeier, C. A., Litchman, E. & Levin, S. A. Phytoplankton growth and stoichiometry under multiple nutrient limitation. *Limnol. Oceanogr.* **49**, 1463–1470 (2004).
8. Falkowski, P. G. Rationalizing elemental ratios in unicellular algae. *J. Phycol.* **16**, 3–6 (2000).
9. Geider, R. J. & La Roche, J. Redfield revisited: variability of C:N:P in marine microalgae and its biochemical basis. *Eur. J. Phycol.* **37**, 1–17 (2002).
10. Klausmeier, C. A., Litchman, E., Daufresne, T. & Levin, S. A. Optimal nitrogen-to-phosphorus stoichiometry of phytoplankton. *Nature* **429**, 171–174 (2004).
11. LaRoche, J. & Breitbarth, E. Importance of diazotrophs as a source of new nitrogen in the ocean. *J. Sea Res.* **53**, 67–91 (2005).
12. Dong, Z., Wang, X. & Liu, L. Wind erosion in semiarid China: An overview. *J. Soil Water Conserv.* **55**, 439–444 (2000).
13. Pahlow, M. & Riebesell, U. Temporal trends in deep ocean Redfield ratios. *Science* **287**, 831–833 (2000).
14. Sarmiento, J. L., Hughes, T. M. C., Stouffer, R. J., Manabe, S. Simulated response of the ocean carbon cycle to anthropogenic climate warming. *Nature* **393**, 245–249 (1998).
15. Arrigo, K. R. *et al.* Phytoplankton community structure and the drawdown of nutrients and CO_2 in the Southern Ocean. *Science* **283**, 365–367 (1999).
16. Smetacek, V. A watery arms race. *Nature* **411**, 745–745 (2001).
17. Gordon, L. I. *et al.* Seasonal evolution of hydrographic properties in the Ross Sea, Antarctica, 1996–1997. *Deep-Sea Res. II* **47**, 3095–3117 (2000).
18. Sarmiento, J. L., Gruber, N., Brzezinski, M. A. & Dunne, J. P. High-latitude controls of thermocline nutrients and low latitude biological productivity. *Nature* **427**, 56–60 (2004).
19. Beman, M., Arrigo, K. R. & Matson, P. A. Agricultural runoff fuels large phytoplankton blooms in vulnerable areas of the ocean. *Nature* **434**, 211–214 (2005).
20. Von Liebig, J. *Chemistry and its Application to Agriculture and Physiology* (Taylor and Walton, London, 1840).
21. Bryant, D. A. The beauty in small things revealed. *Proc. Natl Acad. Sci. USA* **100**, 9647–9649 (2003).
22. Seppala, J., Tamminen, T., & Kaitala, S. Experimental evaluation of nutrient limitation of phytoplankton communities in the Gulf of Riga. *J. Mar. Syst.* **23**, 107–126 (1999).
23. Yin, K., Qian, P.-Y., Chen, J. C., Hsieh, D. P. H. & Harrison, P. J. Dynamics of nutrients and phytoplankton biomass in the Pearl River estuary and adjacent waters of Hong Kong during summer: Preliminary evidence for phosphorus and silicon limitation. *Mar. Ecol. Prog. Ser.* **194**, 295–305 (2000).
24. Mills, M. M., Ridame, C., Davies, M., La Roche, J. & Geider, R. J. Co-limitation of nitrogen fixation by iron and phosphorus in the Eastern North Atlantic. *Nature* **429**, 292–294 (2004).
25. Sunda, W. G. & Huntsman, S. A. Interrelated influence of iron, light and cell size on marine phytoplankton growth. *Nature* **390**, 389–392 (1997).
26. Maldonado, M. T., Boyd, P. W., Harrison, P. J. & Price, N. M. Co-limitation of phytoplankton growth by light and Fe during winter in the NE subarctic Pacific Ocean. *Deep-Sea Res. II* **46**, 2475–2485 (1999).
27. Timmermans, K. R. *et al.* Co-limitation by iron and light of *Chaetoceros brevis*, *C. dicitra* and *C. calcitrans* (Bacillariophyceae). *Mar. Ecol. Prog. Ser.* **217**, 287–297 (2001).
28. Shaked, Y., Xu, Y., Leblanc, K. & Morel, F. M. M. Zinc availability and alkaline phosphatase activity in *Emiliania huxleyi*: Implications for Zn-P co-limitation in the ocean. *Limnol. Oceanogr.* (in the press).
29. Dyhrman, S. T. & Palenik, B. Characterization of ectoenzyme activity and phosphate-regulated proteins in the coccolithophorid *Emiliania huxleyi*. *J. Plank. Res.* **25**, 1215–1225 (2003).
30. Suzumura, M. & Ingall, E. D. Distribution and dynamics of various forms of phosphorus in seawater: Insights from field observations in the Pacific Ocean and a laboratory experiment. *Deep-Sea Res. I* **51**, 1113–1130 (2004).
31. Karl, D. *et al.* The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature* **388**, 533–538 (1997).
32. Wu, J. F., Sunda, W., Boyle, E. A., & Karl, D. M. Phosphate depletion in the western North Atlantic Ocean. *Science* **289**, 759–762 (2000).
33. Niemi, A. Blue-green algal blooms and N:P ratios in the Baltic Sea. *Acta Bot. Fenn.* **110**, 57–61 (1979).
34. Capone, D. G. *et al.* Nitrogen fixation by *Trichodesmium* spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Global Biogeochem. Cycles* **19**, GB2024, 10.1029/2004GB002331 (2005).
35. Michaels, A. F., Bates, N. R., Buesseler, K. O., Carlson, C. A. & Knap, A. H. Carbon-cycle imbalances in the Sargasso Sea. *Nature* **372**, 537–540 (1994).
36. Capone, D. G. & Carpenter, E. J. Nitrogen fixation in the marine environment. *Science* **217**, 1140–1142 (1982).

37. Lipschultz, F. & Owens, N. J. P. An assessment of nitrogen fixation as a source of nitrogen to the North Atlantic Ocean. *Biogeochemistry* **35**, 261–274 (1996).
38. Galloway, J. N., Schlesinger, W. H., Levy, H., Michaels, A. & Schnoor, J. L. Nitrogen fixation: Anthropogenic enhancement—environmental response. *Global Biogeochem. Cycles* **9**, 235–252 (1995).
39. Capone, D. G., Zehr, J. P., Paerl, H. W., Bergman, B. & Carpenter, E. J. *Trichodesmium*, a globally significant marine cyanobacterium. *Science* **276**, 1221–1229 (1997).
40. Ganeshram, R. S., Pedersen, T. F., Calvert, S. E. & Francois, R. Reduced nitrogen fixation in the glacial ocean inferred from changes in marine nitrogen and phosphorus inventories. *Nature* **415**, 156–159 (2002).
41. Karl, D. *et al.* Dinitrogen fixation in the world's oceans. *Biogeochemistry* **57**, 47–98 (2002).
42. Michaels, A., Karl, D. & Capone, D. Element stoichiometry, new production and nitrogen fixation. *Oceanography* **14**, 68–77 (2001).
43. Tyrrell, T. The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature* **400**, 525–538 (1999).
44. Raven, J. A. The iron and molybdenum use efficiencies of plant growth with different energy, carbon and nitrogen sources. *New Phytol.* **109**, 279–287 (1988).
45. Falkowski, P. G. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO₂ in the ocean. *Nature* **387**, 272–275 (1997).
46. Sanudo-Wilhelmy, S. A. *et al.* Phosphorus limitation of nitrogen fixation by *Trichodesmium* in the central Atlantic Ocean. *Nature* **411**, 66–69 (2001).
47. Flynn, K. J. & Hipkin, C. R. Interactions between iron, light, ammonium, and nitrate: Insights from the construction of a dynamic model of algal physiology. *J. Phycol.* **35**, 1171–1190 (1999).
48. Voss, M., Croot, P., Lochte, K., Mills, M. & Peeken, I. Patterns of nitrogen fixation along 10°N in the tropical Atlantic. *Geophys. Res. Lett.* **31**, L23S09 (2004).
49. Shriadah, M. A., Okbah, M. A., & El-Deek, M. S. Trace metals in the water columns of the red sea and the gulf of Aqaba, Egypt. *Wat. Air Soil Pollut.* **153**, 115–124 (2004).
50. Carpenter, E. J., Subramaniam, A. & Capone, D. G. Biomass and primary productivity of the cyanobacterium, *Trichodesmium* spp., in the southwestern tropical N. Atlantic Ocean. *Deep-Sea Res.* **51**, 173–203 (2004).
51. Venrick, E. L. The distribution and significance of *Richelia intracellularis* Schmidt in the North Pacific central gyre. *Limnol. Oceanogr.* **19**, 437–444 (1974).
52. Carpenter, E. J. *et al.* Extensive bloom of a N₂ fixing symbiotic association in the tropical Atlantic Ocean. *Mar. Ecol. Prog. Ser.* **188**, 273–283 (1999).
53. Zehr, J. P. *et al.* Unicellular cyanobacteria fix N₂ in the subtropical North Pacific Ocean. *Nature* **412**, 635–638 (2001).
54. Zehr, J. P. & Ward, B. B. Nitrogen cycling in the ocean: New perspectives on processes and paradigms. *Appl. Env. Microbiol.* **68**, 1015–1024 (2002).
55. Church, M. J., Jenkins, B. D., Karl, D. M. & Zehr, J. P. Vertical distributions of nitrogen-fixing phylotypes at Stn ALOHA in the oligotrophic North Pacific Ocean. *Aquat. Microb. Ecol.* **38**, 3–14 (2005).
56. Montoya, J. P. *et al.* High rates of N₂ fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature* **430**, 1027–1031 (2004).
57. Hansell, D. A., Bates, N. R. & Olson, D. B. Excess nitrate and nitrogen fixation in the North Atlantic Ocean. *Mar. Chem.* **84**, 243–265 (2004).
58. Dore, J. E., Brum, J. R., Tupas, L. M. & Karl, D. M. Seasonal and interannual variability in sources of nitrogen supporting export in the oligotrophic subtropical North Pacific Ocean. *Limnol. Oceanogr.* **47**, 1595–1607 (2002).
59. Codispoti, L. A. *et al.* The oceanic fixed nitrogen and nitrous oxide budgets: Moving targets as we enter the anthropocene? *Scientia Marina* **65** (Suppl. 2), 85–105 (2001).
60. Richards, F. A. in *Chemical Oceanography* (eds Ripley, J. P. & Skirrow, G.) 611–645 (Academic, London, 1965).
61. Mulder, A., van de Graaf, A. A., Robertson, L. A. & Kuenen, J. G. Anaerobic ammonium oxidation discovered in a denitrifying fluidized bed reactor. *FEMS Microbiol. Ecol.* **16**, 177–184 (1995).
62. Van de Graaf, A. A. *et al.* Anaerobic oxidation of ammonium is a biologically mediated process. *Appl. Environ. Microbiol.* **61**, 1246–1251 (1995).
63. Thamdrup, B. & Dalsgaard, T. Production of N₂ through anaerobic ammonium oxidation coupled to nitrate reduction in marine sediments. *Appl. Environ. Microbiol.* **68**, 1312–1318 (2002).
64. Strous, M. *et al.* Missing lithotroph identified as new planctomycete. *Nature* **400**, 446–449 (1999).
65. Dalsgaard, T., Canfield, D. E., Petersen, J., Thamdrup, B. & Acuña-González, J. Anammox is a significant pathway of N₂ production in the anoxic water column of Golfo Dulce, Costa Rica. *Nature* **422**, 606–608 (2003).
66. Sinnighe Damsté, J. S. *et al.* Linearly concatenated cyclobutane (ladderane) lipids from a dense bacterial membrane. *Nature* **419**, 708–712 (2002).
67. Schmid, M. *et al.* Molecular evidence for genus level diversity of bacteria capable of catalyzing anaerobic ammonium oxidation. *Syst. Appl. Microbiol.* **23**, 93–106 (2000).
68. Rysgaard, S. & Glud, R. N. Anaerobic N₂ production in Arctic sea ice. *Limnol. Oceanogr.* **49**, 86–94 (2004).
69. Kuypers, M. M. M. *et al.* Massive nitrogen loss from the Benguela upwelling system through anaerobic ammonium oxidation. *Proc. Natl Acad. Sci. USA* **102**, 6478–6483 (2005).
70. Carr, M. E. Estimation of potential productivity in Eastern Boundary Currents using remote sensing. *Deep-Sea Res.* **49**, 9–80 (2002).
71. Tyrrell, T. & Lucas, M. I. Geochemical evidence of denitrification in the Benguela upwelling system. *Cont. Shelf Res.* **22**, 2497–2511 (2002).
72. Devol, A. H. Solution to a marine mystery. *Nature* **422**, 575–576 (2003).
73. Service, R. F. Oceanography—New dead zone off Oregon coast hints at sea change in currents. *Science* **305**, 1099 (2004).
74. Mohan, S. B., Schmid, M., Jetten, M. & Cole, J. Detection and widespread distribution of the *nrfA* gene encoding nitrite reduction to ammonia, a short circuit in the biological nitrogen cycle that competes with denitrification. *FEMS Microbiol. Ecol.* **49**, 433–443 (2004).

Acknowledgements I would like to thank S. Rysgaard, T. Dalsgaard, and M. Kuypers for assistance with the anammox section of this manuscript. D. Capone, E. Carpenter, and A. Subramaniam provided valuable insight about N₂ fixation. I would like to thank M. Davey and G. Tarran for providing their unpublished flow cytometry data from their North Atlantic work. Thanks also to M. Mills, G. van Dijken, A. Tagliabue, R. Labiosa, T. Reddy, S. Pabi, L. Kropuenske and B. Saenz for comments on earlier drafts of this manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.R.A. (arrigo@stanford.edu).

Viruses in the sea

Curtis A. Suttle¹

Viruses exist wherever life is found. They are a major cause of mortality, a driver of global geochemical cycles and a reservoir of the greatest genetic diversity on Earth. In the oceans, viruses probably infect all living things, from bacteria to whales. They affect the form of available nutrients and the termination of algal blooms. Viruses can move between marine and terrestrial reservoirs, raising the spectre of emerging pathogens. Our understanding of the effect of viruses on global systems and processes continues to unfold, overthrowing the idea that viruses and virus-mediated processes are sidebars to global processes.

For years viruses were known to exist in seawater, but reports 15 years ago caused great excitement by demonstrating not only that viruses are abundant, but that they infect the dominant organisms in the ocean^{1–3}. These observations occurred against the backdrop of a major shift in thinking among oceanographers to acknowledge that bacteria and microbial processes are important players in the oceans. For example, because viruses are significant agents of microbial mortality, they have an effect on nutrient cycling^{4–6}. Moreover, the narrow host range of most viruses suggests that infection is important in controlling the composition of planktonic communities^{6–8}. Interest in the vast viral communities in the sea continues to expand as the relevance of viruses to evolution, pathogen emergence and even exobiology begins to be explored.

Several excellent reviews^{4–7} have captured the advances in our understanding of marine viruses and their role in the ocean. This review focuses on areas where our knowledge is changing rapidly, where methodological problems have impeded progress and where new data are altering perceptions. Without doubt, viruses are the most abundant and genetically diverse 'life forms' in the ocean. They are major pathogens of planktonic organisms and consequently are significant players in nutrient and energy cycling. As well, they are pathogens of higher organisms and there is good evidence that some viruses move between marine and terrestrial reservoirs. Recognition that viruses play a major role in marine ecosystems has added a significant new dimension to our understanding of biological oceanographic processes.

Total virus abundance has been underestimated

Viruses are extremely abundant in aquatic systems. The first observations by transmission electron microscopy (TEM)^{1,2} indicated that, typically, there are $\sim 10^7$ viruses ml^{-1} and that abundance decreases with depth and distance from the shore^{9,10}. In general, abundance correlates with system productivity and is highest where bacteria and chlorophyll are greatest^{11,12}. In marine sediments, abundances are even higher, with 10^8 – 10^9 viruses cm^{-3} typical in nearshore surface sediments^{1,10,13}. Even 100 m below the sediment surface¹⁵ viruses can be plentiful, although at the sediment surface in the deep ocean they seem to be less abundant¹⁶.

Epifluorescence microscopy (EM) is now the preferred method for counting viruses because of its higher accuracy and precision^{17,18}, although flow cytometry shows promise as a high-throughput method¹⁹. The shift to EM-based techniques has not been without problems, including significant differences in results among methodologies²⁰, concerns about reproducibility and effects of sample storage^{19,21}. For instance, estimates of viral abundance performed using samples not immediately processed or frozen in liquid nitrogen can be an order of

magnitude too low. Currently, our best estimates range from $\sim 3 \times 10^6$ viruses ml^{-1} in the deep sea^{22,23} to $\sim 10^8$ viruses ml^{-1} in productive coastal waters. Assuming the volume of the oceans is 1.3×10^{21} l and the average abundance of viruses is 3×10^9 l^{-1} , then ocean waters contain $\sim 4 \times 10^{30}$ viruses. Because a marine virus contains about 0.2 fg of carbon and is about 100 nm long, this translates into 200 Mt of carbon in marine viruses. If the viruses were stretched end to end they would span ~ 10 million light years. In context, this is equivalent to the carbon in ~ 75 million blue whales ($\sim 10\%$ carbon, by weight²⁴), and is ~ 100 times the distance across our own galaxy. This makes viruses the most abundant biological entity in the water column of the world's oceans, and the second largest component of biomass after prokaryotes.

But total viral abundance alone does not give us an indication of infectivity. Most viruses in seawater seem to be infectious²⁵, and some can remain infectious in sediments for long periods, from decades to a hundred years or more^{26,27}. Estimating the abundance of infectious viruses is complicated by strain specificity; yet in offshore waters most collisions between bacteria and viruses seem to result in infection, suggesting that selection for resistance is low²⁸. Even so, viruses infecting specific hosts can be extremely abundant. For example, viruses infecting single strains of the cyanobacterium *Synechococcus*²⁸ or the photosynthetic flagellate *Micromonas pusilla*²⁹ can occur in excess of 10^5 infectious units ml^{-1} . Yet, even the most permissive hosts are not sensitive to infection by all viruses that infect a given species; therefore, even these are underestimates. Ultimately, high strain specificity combined with poor representation in culture of the dominant microbes in the sea means that the absolute abundance of infectious viruses must be deduced by inference.

Unexplored diversity

Virus form provides insight into function

Not only are viruses abundant in oceans but, as is becoming clear, they also harbour enormous genetic and biological diversity. TEM studies of marine viral communities¹ and phage isolates³⁰ reveal a plethora of morphotypes, whereas host-range studies show complex patterns of resistance and susceptibility³¹.

Among marine phages (Fig. 1), those with contractile tails (such as myoviruses and T4-like viruses) and long flexible tails (such as siphoviruses and lambda-like viruses) are most frequently isolated^{32–34}, even though TEM suggests that phages with short non-contractile tails (such as podoviruses and T7-like viruses) and without tails are most abundant. Because 'lifestyles' among tailed phages differ, morphology provides clues about host range and viral replication. For example, myoviruses are typically lytic and often have a broader host range than other tailed phages, even infecting different species of bacteria^{33,34}. By

¹Department of Chemistry, University of California, Berkeley and the Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

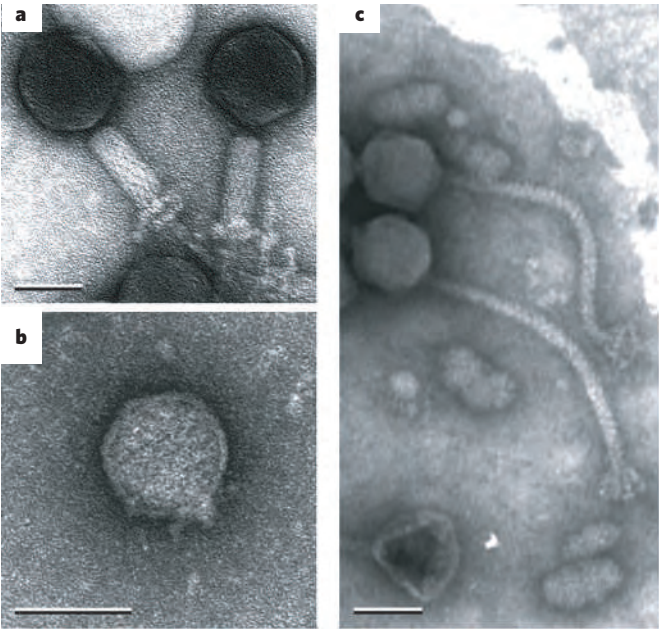


Figure 1 | The three families of tailed dsDNA viruses (phages) that infect bacteria. **a**, Myoviruses are often the most commonly isolated phage from natural marine viral communities. They have contractile tails, are typically lytic and often have relatively broad host ranges. **b**, Podoviruses have a short non-contractile tail, are also typically lytic and have very narrow host ranges. They are less commonly isolated from seawater. **c**, Siphoviruses have long non-contractile tails. They are frequently isolated from seawater, often have a relatively broad host range, and many are capable of integrating into the host genome. Scale bar, 50 nm.

contrast, the host range of podoviruses is generally narrowest, with siphoviruses being intermediate. However, many siphoviruses can integrate into the host genome and be passed from generation to generation.

Morphotype provides some insights into the selective pressures facing virus communities. Myoviruses with their broader host ranges can quickly take advantage of increases in host populations, consistent with *r*-selection (short generation times and high reproductive rates). By contrast, many siphoviruses can archive their genomes in host cells, tying their replication rate to that of the host, until an environmental cue triggers the lytic cycle. This suggests that siphoviruses are more *K*-selected (longer generation times and lower reproductive rates). There is much to learn about the biological and genetic diversity of marine phages. In culture, the phages used are unlikely to be representative of the dominant phages in the ocean, because the most abundant prokaryotes have proven very difficult to grow in culture,

although this is beginning to change.

The story with respect to viruses infecting eukaryotes is even more complex. The first virus isolates infecting eukaryotic phytoplankton belonged to a group of large double-stranded (ds) DNA viruses, the *Phycodnaviridae*, and included viruses that infected important taxa of marine primary producers including toxic bloom formers and macroalgae^{26,35}. The scene is changing rapidly as the isolation of many previously unknown viruses (Table 1) greatly enriches the taxonomic and phylogenetic space of known viral 'life'. For instance, a single-stranded (ss)RNA virus that infects a toxic bloom-forming alga (*Heterosigma akashiwo*) led to the creation of the *Marnaviridae*³⁶, probably the first of many new marine virus families. Other examples include a previously unknown dsRNA virus that infects the photosynthetic flagellate, *Micromonas pusilla*³⁷, a ssRNA virus that infects a thraustochytrid fungus³⁸, and a nuclear-inclusion virus (NIV) that has both ss and dsDNA and infects the diatom alga *Chaetoceros salsugineum*³⁹, and has virtually no similarity to known viruses. Other NIVs that infect *H. akashiwo*⁴⁰ (Fig. 2) and *Cheateoceros*⁴¹, as well as a large dsDNA virus that infects a marine heterotrophic protist⁴², remain to be characterized. The largest virus genome belongs to Mimivirus⁴³, which also infects a protist, whereas viruses that cause disease in shrimp⁴⁴ and crabs⁴⁵ are distantly related to other viruses, suggesting that heterotrophs will also yield an untapped oasis of unexplored viral diversity.

Viral genetic diversity is extremely high

Culture-independent approaches indicate that we are just scraping the surface of viral life in the oceans. No gene is found universally in viruses but some genes are representative of specific subsets of the viral community. The first studies used the DNA polymerase genes of *Phycodnaviridae*⁴⁶ to reveal enormous genetic variation that was not represented in cultures, and showed that very similar sequences were present in distant oceans⁴⁷. Even more striking results were obtained in studies targeting a subset of myoviruses: tremendous diversity occurred on large⁴⁸ and small⁴⁹ spatial scales, with most of the sequences falling into groups with no cultured representatives^{50,51}. Even more remarkable is that nearly identical sequences at the nucleotide level occurred in environments as far-reaching as the Southern Ocean, the Gulf of Mexico and a melt-water pond on an Arctic ice shelf⁵¹. Similarly, a study that targeted podoviruses found that indistinguishable sequences were ubiquitous in the environment⁵². These results indicate that in phycoviruses, podoviruses and myoviruses, and by inference probably other groups, co-infection results in viral genes being sequentially passed in many small 'steps' through a series of organisms. This emphasizes that, from a virocentric perspective, bacteria and protists operate as vehicles for viral 'sex', allowing viral genes to spread widely. This remarkable diversity is not confined to dsDNA viruses infecting bacteria and protists. A study

Table 1 | Some unusual aquatic viruses whose genomic sequences have recently been completed

Virus	Family of proteins	Nucleic acid	Genome size (bp)	Number
<i>Chaetoceros salsugineum</i> nuclear inclusion virus ³⁹	Unassigned	ssDNA / dsDNA	6,005 ssDNA 997 dsDNA	Unknown
<i>Emiliania huxleyi</i> virus86 (ref. 63)	<i>Phycodnaviridae</i>	dsDNA	407,339	472
<i>Heterosigma akashiwo</i> RNA virus ³⁶	<i>Marnaviridae</i>	ssRNA	8,600	6 or 7
<i>Micromonas pusilla</i> ³⁷	<i>Reoviridae</i>	dsRNA	26,000 on 11 segments	Unknown
<i>Ectocarpus siliculosus</i> virus ⁸⁵	<i>Phycodnaviridae</i>	dsDNA	335,593	240
Mimivirus ⁴³	<i>Mimiviridae</i>	dsDNA	1,181,404	911
White spot syndrome virus ⁴⁴	<i>Nimaviridae</i>	dsDNA	305,107	531

All infect eukaryotic phytoplankton with the exception of the *Ectocarpus*, Mimivirus, and white-spot viruses, which infect a brown alga, freshwater protist and penaeid shrimp, respectively.

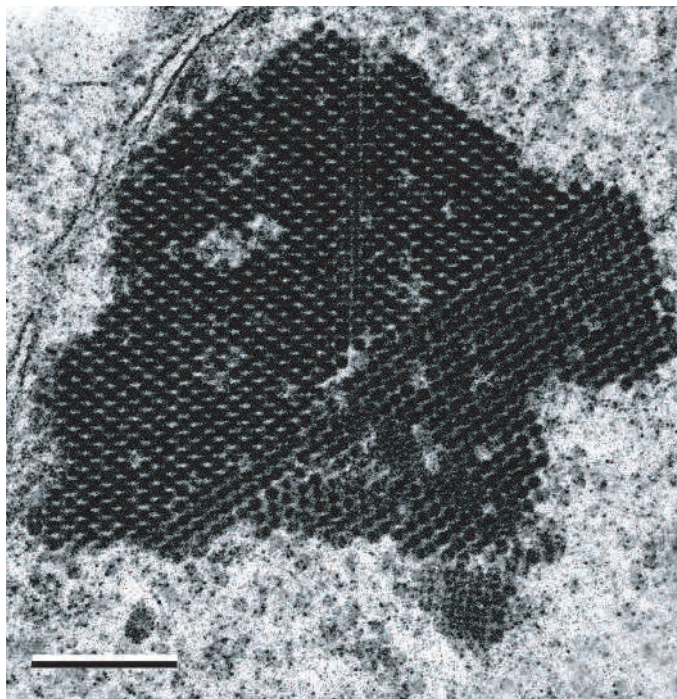


Figure 2 | The nuclear inclusion virus of *Heterosigma akashiwo*. There are a large number of viruses infecting marine protists that have only partially been characterized. Shown is a nuclear inclusion virus that infects the toxic bloom-forming alga *H. akashiwo*. Scale bar, 250 nm.

that targeted the RNA-dependent RNA polymerase of picorna-like viruses found a putative four new families in a few seawater samples⁵³. This is particularly remarkable given the few existing families of known picorna-like viruses.

The extent of viral diversity in the sea is driven home by metagenomic studies in coastal waters⁵⁴ and sediments⁵⁵ that show there are perhaps several thousand viral genotypes in 200 l of seawater and a million in 1 kg of sediment⁵⁶. The communities have an uneven distribution, with the most abundant genotypes making up less than 5% of the communities, whereas the majority of genotypes comprise < 0.01 of the communities^{56,57}. The genetic richness of the communities is revealed by the fact that 60–80% of the sequences were not similar (*E*-value > 0.001) to those in databases⁵⁷. By contrast, ~90% of putative genes from metagenomic data for prokaryotic communities had recognizable similarity to database sequences⁵⁷. Clearly, metagenomic data indicate that marine viral communities contain much greater genetic richness than their prokaryotic counterparts, and are much less sampled.

Host cells are vessels for viral sex

Sequencing of virus isolates also reveals how little we know. The first marine phages to be sequenced infect *Roseobacter*⁵⁸, *Vibrio parahaemolyticus*⁵⁹ and *Synechococcus* sp.⁶⁰. The results showed that these viruses appear to be ancestral to their terrestrial relatives; however, most of the putative genes had no significant similarity to those in databases. Subsequently, other marine phages have been sequenced^{61,62}, revealing that most of the putative genes code for proteins of unknown function, although some are homologous to genes of surprising metabolic function, such as those associated with photosynthesis, carbon metabolism and phosphate stress⁶². Similarly, analysis of the first viral genome to be completely sequenced that infects eukaryotic phytoplankton revealed that of 472 putative coding sequences, only 14% have significant similarity to database sequences⁶³.

One of the most surprising observations to stem from genomic analysis of cyanophages was that they contain homologues to *psbA* and

psbD genes, which encode key components of photosynthesis⁶⁴. At least one of these, and a varying number of other related genes, occur in phages that infect *Synechococcus*⁶⁵, *Prochlorococcus*⁶⁶ or both. These genes are widespread, with most marine cyanophage isolates carrying *psbA*⁶⁵. Sequence analysis reveals that lateral gene transfer has occurred, probably numerous times, between cyanobacteria and their viruses⁶⁷; however, these exchanges appear to be infrequent, particularly with respect to *Synechococcus*. The result is that photosynthesis genes in cyanophages have their own evolutionary history, with phage *psbA* genes forming distinct monophyletic groups^{67,68}. This example demonstrates that viruses capture genes of host origin and exchange them among viral progeny, resulting in photosynthetic genes that are now clearly viral.

Accurate estimates of virus-mediated mortality are elusive

At least a half-dozen approaches have been used to infer the impact of viruses on microbial mortality. All of them suffer from poorly constrained assumptions. The elegant approach of inferring mortality from the percentage of visibly infected cells² requires estimates of the proportion of the lytic cycle during which cells are visibly infected. Few data exist for marine microbes and those that do are highly variable. Extrapolations from increasing the abundance of viruses only demonstrate the potential effect of viral infection³. Inferring mortality from decay rates of virus communities⁶⁹ or viral tracers⁷⁰ assumes that viral production and removal are balanced, that tracers are representative of *in situ* communities and that the number of viruses produced per lytic event (burst size) is known. Calculations on the basis of fluorescent-virus tracers⁷¹ also require an estimate of burst size, and this method is not amenable to use in sunlight. Extrapolation from synthesis rates of viral DNA⁷² requires conversion factors and is sensitive to contamination from bacterial DNA. Converting from directly measured rates of viral production to microbial mortality rates⁷³ requires an estimate of burst size and considerable sample manipulation. Finally, measurements based on changes in net growth rate as a function of viral dilution⁷⁴ require sample manipulation and assumptions about the relationship between infection and host-cell mortality.

Although it is discouraging that no method gives high precision and accurate estimates of virus-mediated mortality, it is encouraging that these diverse approaches, many of which rely on assumptions that are independent of each other, consistently indicate that viruses cause significant microbial mortality in a wide range of environments. Yet, many environments remain inadequately sampled. This includes the

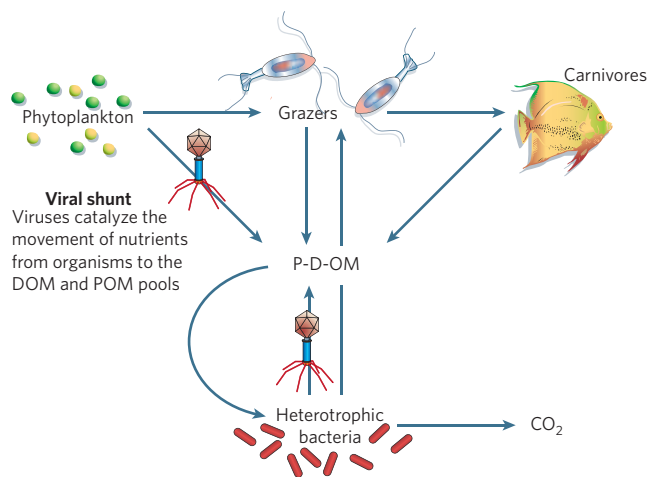


Figure 3 | Viruses are catalysts for biogeochemical cycling. Viruses short-circuit the flow of carbon and nutrients from phytoplankton and bacteria to higher trophic levels by causing the lysis of cells and shunting the flux to the pool of dissolved and particulate organic matter (D-P-OM). The result is that more of the carbon is respired, thereby decreasing the trophic transfer efficiency of nutrients and energy through the marine foodweb.

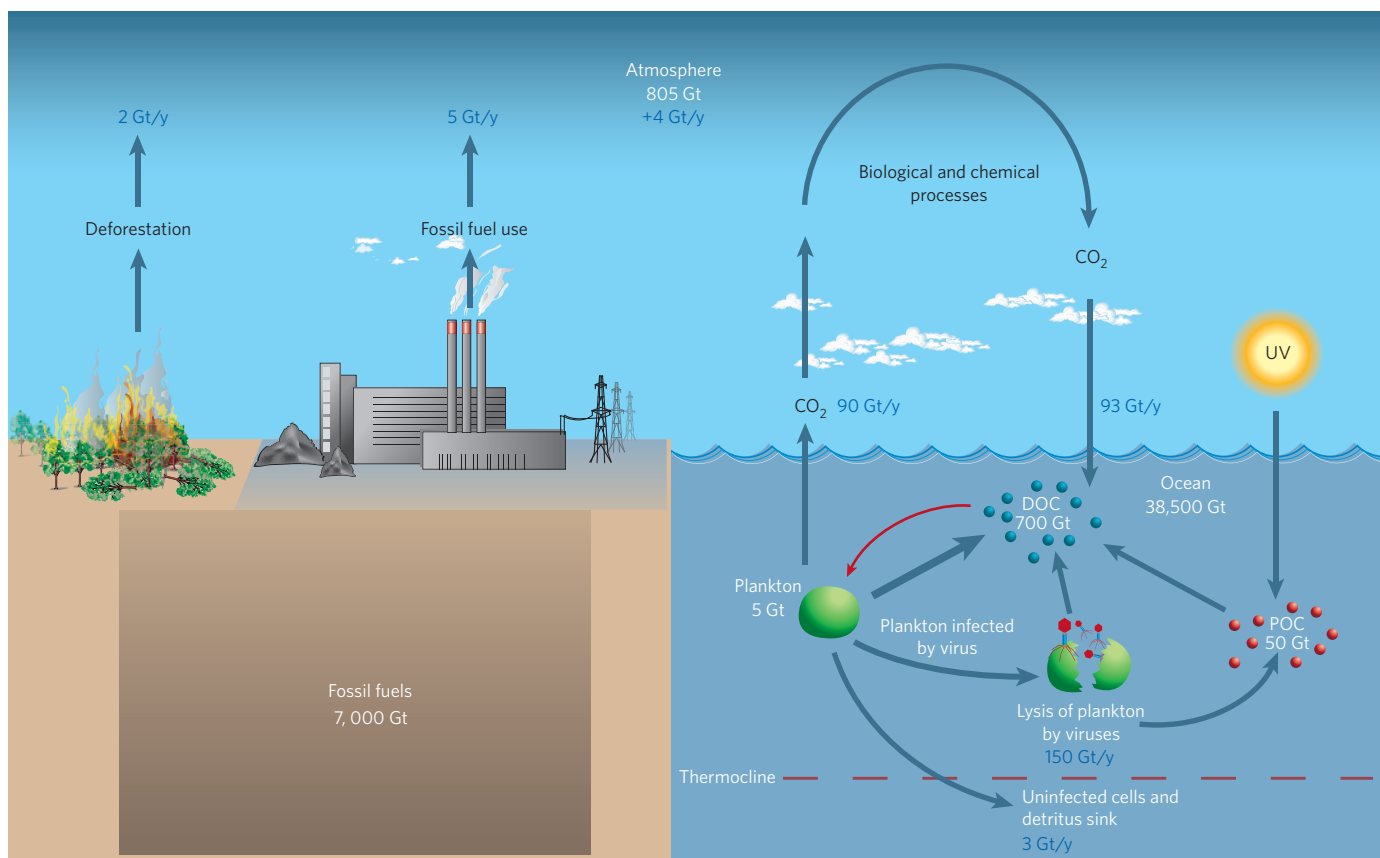


Figure 4 | Viruses can affect the efficiency of the biological pump. Viruses cause the lysis of cells, converting them into particulate organic carbon (POC) and dissolved organic carbon (DOC). This reduces the rate at which C sinks from the surface layer into the deep ocean where the carbon is trapped for millennia (biological pump). Instead the carbon is retained in the surface waters where it is photo-oxidized and respired, in chemical equilibrium with the atmosphere. The net effect is a faster rate of CO_2 build-up in the atmosphere than would occur if the POC were 'exported' to the deep ocean.

oligotrophic ocean, where slow rate processes make it difficult to obtain reliable data except from visibly infected cells or viral decay rates. Moreover, the assumptions for calculating mortality rates are particularly poorly grounded for open-ocean species. Even within relatively productive environments, estimates of the contribution of viruses to total mortality range from undetectable to 100%. In many cases, the wide range in estimates is probably real, and reflects differences between locations and times. Nonetheless, accurate estimates of virus-mediated mortality remain elusive, and we are not much further ahead than a decade ago when viruses were estimated to kill ~20–40 % of marine bacteria on a daily basis⁷⁵ and contribute to microbial mortality at a level similar to that of grazing by zooplankton⁷⁶.

Viruses are catalysts of global nutrient cycles

Given that viruses cause a significant, albeit variable, amount of marine microbial mortality, it implies that they also play an important role in marine geochemical cycles. Simple models^{4,5} and model systems⁷⁷ demonstrate that viruses are catalysts that accelerate the transformation of nutrients from particulate (living organisms) to dissolved states, where it can be incorporated by microbial communities (Fig. 3). A net effect of this shunt is to increase community respiration and decrease the efficiency of carbon transfer to higher trophic levels. In addition, cell lysis converts particulate organic carbon (POC) into dissolved (therefore lower levels of cellular carbon) sinks (Fig. 4), resulting in more carbon being respired in the surface waters. This is significant for global carbon cycling because sinking of POC results in the net transfer of about 3 Gt of carbon between near-surface and deep waters thus the build-up of CO_2 in the atmosphere is only about half of what it otherwise would be. An exception occurs in some phytoplank-

ton, where virus-infected cells sink rapidly⁷⁸, potentially increasing the transport of cells to deeper waters²⁷. Nutrients other than carbon are also released by viral lysis^{79,80}. As these nutrients are largely organically bound, this can affect their availability and pathways of cycling. In some cases, released nutrients such as iron can fill a major portion of the requirements of other organisms⁸⁰. As well, the small size of viruses makes them excellent nucleation sites for mineralization of iron and perhaps other metals⁸¹.

Marine viruses and disease

Viruses are not only players in microbial mortality and geochemical cycling, they are also progenitors of disease in higher organisms. Our limited knowledge of viral diseases in non-microbial marine organisms stems almost entirely from effects on fisheries from obvious instances of visible disease or large mortality events. Although viruses infect marine organisms ranging from crustaceans to whales, we know little about modes of infection and transmission, or the reservoirs of these viruses in nature. Some of these viruses pose potential health risks to humans. For example, calici and distemper viruses are thought to cycle between marine and terrestrial mammals, and some marine caliciviruses are thought to cause disease in humans^{82,83}. Similarly, there is evidence that marine birds harbour avian flu, particularly the dangerous H5N1 strain⁸⁴. We have very little understanding of the natural reservoirs of viruses that are carried by, or cause disease in, marine animals and know even less about their potential to spread to terrestrial systems. Our emerging knowledge of the enormous diversity of viruses in the marine milieu suggests that the oceans are potential reservoirs of many unknown causative agents of disease.

A look to the future

Our understanding of viruses and virus-mediated processes in the oceans is developing rapidly. From a very few studies in the early 1990s indicating that viruses were abundant and active in the sea, we have reached a point where it is clear that viruses constitute the greatest genetic diversity in the ocean, are important agents of mortality and major players in global geochemical cycles. Yet, for the most part, the depth of understanding is extremely limited in all of these areas. We have not determined the genetic richness of the marine viral metagenome within an order of magnitude, estimates of mortality are highly variable and poorly constrained, and the ways in which viruses influence global geochemical cycles are only beginning to be elucidated. Moreover, the role of marine viral communities in emerging and established diseases in marine and terrestrial ecosystems and the cycling of viruses between these reservoirs is largely unknown. Knowledge of marine viruses and their role in the global ecosystem will influence the spectrum of our thinking, ranging from quantitative models of global geochemical cycling to the localization of aquaculture facilities. Few would have predicted that the observation a decade and a half ago of high viral abundances in seawater^{1,2} would have had such a profound influence on our understanding of biological oceanographic processes, evolution and geochemical cycling. ■

- Bergh, O., Børsheim, K. Y., Bratbak, G. & Haldal, M. High abundance of viruses found in aquatic environments. *Nature* **340**, 467–468 (1989).
- Proctor, L. M. & Fuhrman, J. A. Viral mortality of marine bacteria and cyanobacteria. *Nature* **343**, 60–62 (1990).
- Suttle, C. A., Chan, A. M. & Cottrell, M. T. Infection of phytoplankton by viruses and reduction of primary productivity. *Nature* **347**, 467–469 (1990).
- Wilhelm, S. W. & Suttle, C. A. Viruses and nutrient cycles in the sea. *Bioscience* **49**, 781–788 (1999).
- Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541–548 (1999).
- Wommack, K. E. & Colwell, R. R. Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **64**, 69–114 (2000).
- Weinbauer, M. G. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* **28**, 127–181 (2004).
- Weinbauer, M. G. & Rassoulzadegan, F. Are viruses driving microbial diversification and diversity? *Environ. Microbiol.* **6**, 1–11 (2004).
- Cochlan, W. P., Wikner, J., Steward, G. F., Smith, D. C. & Azam, F. Spatial distribution of viruses, bacteria and chlorophyll *a* in neritic, oceanic and estuarine environments. *Mar. Ecol. Prog. Ser.* **92**, 77–87 (1993).
- Paul, J. H., Rose, J. B., Jiang, S. C., Kellogg, C. A. & Dickson, L. Distribution of viral abundance in the reef environment of Key Largo, Florida. *Appl. Environ. Microbiol.* **59**, 718–724 (1993).
- Boehme, J. *et al.* Viruses, bacterioplankton, and phytoplankton in the southeastern Gulf of Mexico: distribution and contribution to oceanic DNA pools. *Mar. Ecol. Prog. Ser.* **97**, 1–10 (1993).
- Maranger, R. & Bird, D. F. Viral abundances in aquatic systems: a comparison between marine and fresh waters. *Mar. Ecol. Prog. Ser.* **121**, 217–226 (1995).
- Hewson, I., O'Neil, J. M., Fuhrman, J. A. & Dennison, W. C. Virus-like particle distribution and abundance in sediments and overlying waters along eutrophication gradients in two subtropical estuaries. *Limnol. Oceanogr.* **46**, 1734–1746 (2001).
- Middelboe, M., Glud, R. N. & Finster, K. Distribution of viruses and bacteria in relation to diagenetic activity in an estuarine sediment. *Limnol. Oceanogr.* **48**, 1447–1456 (2003).
- Bird, D. F. *et al.* Subsurface viruses and bacteria in Holocene/Late Pleistocene sediments of Saanich Inlet, BC: ODP holes 1033b and 1034b, Leg 169s. *Mar. Geol.* **174**, 227–239 (2001).
- Danovaro, R., Corinaldesi, C., Dell'anno, A., Fabiano, M. & Corselli, C. Viruses, prokaryotes and DNA in the sediments of a deep-hypersaline anoxic basin (DHAB) of the Mediterranean Sea. *Environ. Microbiol.* **7**, 586–592 (2005).
- Hennes, K. P. & Suttle, C. A. Direct counts of viruses in natural waters and laboratory cultures by epifluorescence microscopy. *Limnol. Oceanogr.* **40**, 1050–1055 (1995).
- Noble, R. T. & Fuhrman, J. A. Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquat. Microb. Ecol.* **14**, 113–118 (1998).
- Brussaard, C. P. D. Optimization of procedures for counting viruses by flow cytometry. *Appl. Environ. Microbiol.* **70**, 1506–1513 (2004).
- Bettarel, Y., Sime-Ngando, T., Amblard, C. & Laveran, H. A comparison of methods for counting viruses in aquatic systems. *Appl. Environ. Microbiol.* **66**, 2283–2289 (2000).
- Wen, K., Ortmann, A. C. & Suttle, C. A. Accurate estimation of viral abundance by epifluorescence microscopy. *Appl. Environ. Microbiol.* **70**, 3862–3867 (2004).
- Guixa-Boixereu, N., Vaque, D., Gasol, J. M., Sanchez-Camara, J. & Pedros-Alio, C. Viral distribution and activity in Antarctic waters. *Deep-Sea Res. II* **49**, 827–845 (2002).
- Ortmann, A. C. & Suttle, C. A. High abundances of viruses in a deep-sea hydrothermal vent system indicates viral mediated microbial mortality. *Deep-Sea Res.* **52**, 1515–1527 (2005).
- Jelmert, A. & Oppen-Berntsen, D. O. Whaling and deep-sea biodiversity. *Conserv. Biol.* **10**, 653–654 (1996).
- Wilhelm, S. W., Weinbauer, M. G., Suttle, C. A., Pledger, R. J. & Mitchell, D. L. Measurements of DNA damage and photoreactivation imply that most viruses in marine surface waters are infective. *Aquat. Microb. Ecol.* **14**, 215–222 (1998).
- Suttle, C. A. in *Viral Ecology* (ed. Hurst, C. J.) 248–286 (Academic, New York, 2000).
- Lawrence, J. E., Chan, A. M. & Suttle, C. A. Viruses causing lysis of the toxic bloom-forming alga, *Heterosigma akashiwo* (Raphidophyceae), are widespread in coastal sediments of British Columbia, Canada. *Limnol. Oceanogr.* **47**, 545–550 (2002).
- Suttle, C. A. & Chan, A. M. Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Appl. Environ. Microbiol.* **60**, 3167–3174 (1994).
- Cottrell, M. T. & Suttle, C. A. Dynamics of a lytic virus infecting the photosynthetic marine picoflagellate, *Micromonas pusilla*. *Limnol. Oceanogr.* **40**, 730–739 (1995).
- Frank, H. & Moebus, K. An electron microscopic study of bacteriophages from marine waters. *Helgolander Meeresunters* **41**, 385–414 (1987).
- Moebus, K. & Nattkemper, H. Bacteriophage sensitivity patterns among bacteria isolated from marine waters. *Helgolander Meeresunters* **34**, 375–385 (1981).
- Kellogg, C. A., Rose, J. B., Jiang, S. C., Thurmond, J. & Paul, J. H. Genetic diversity of related vibriophages isolated from marine environments around Florida and Hawaii, USA. *Mar. Ecol. Prog. Ser.* **120**, 89–98 (1995).
- Wichels, A. *et al.* Bacteriophage diversity in the North Sea. *Appl. Environ. Microbiol.* **64**, 4128–4133 (1998).
- Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**, 1047–1051 (2003).
- Van Etten, J. L., Graves, M. V., Muller, D. G., Boland, W. & Delaroque, N. *Phycodnaviridae* — Large DNA algal viruses. *Arch. Virol.* **147**, 1479–1516 (2002).
- Lang, A. S., Culley, A. I. & Suttle, C. A. Genome sequence and characterization of a virus (HaRNAV) related to picorna-like viruses that infects the marine toxic bloom-forming alga *Heterosigma akashiwo*. *Virology* **320**, 206–217 (2004).
- Brussaard, C. P. D., Noordeloos, A. A. M., Sandaa, R. A., Haldal, M. & Bratbak, G. Discovery of a dsRNA virus infecting the marine photosynthetic protist *Micromonas pusilla*. *Virology* **319**, 280–291 (2004).
- Takao, Y., Nagasaki, K., Mise, K., Okuno, T. & Honda, D. Isolation and characterization of a novel single-stranded RNA virus infectious to a marine fungoid protist, *Schizochytrium* sp. (Thraustochytriaceae, Labyrinthulea). *Appl. Environ. Microbiol.* **71**, 4516–4522 (2005).
- Nagasaki, K. *et al.* Previously unknown virus infects marine diatom. *Appl. Environ. Microbiol.* **71**, 3528–3535 (2005).
- Lawrence, J. E., Chan, A. M. & Suttle, C. A. A novel virus (HaNIV) causes lysis of the toxic bloom-forming alga *Heterosigma akashiwo* (Raphidophyceae). *J. Phycol.* **37**, 216–222 (2001).
- Bettarel, Y. *et al.* Isolation and characterisation of a small nuclear inclusion virus infecting the diatom *Chaetoceros cf. gracilis*. *Aquat. Microb. Ecol.* (in the press).
- Garza, D. R. & Suttle, C. A. Large double-stranded DNA viruses which cause the lysis of marine heterotrophic nanoflagellates (*Bodo* sp.) occur in natural marine virus communities. *Aquat. Microb. Ecol.* **9**, 203–210 (1995).
- Raoult, D. *et al.* The 1.2-megabase genome sequence of mimivirus. *Science* **306**, 1344–1350 (2004).
- Van Hulten, M. C. W. *et al.* The White Spot Syndrome Virus DNA genome sequence. *Virology* **286**, 7–22 (2001).
- Zhang, S., Shi, Z., Zhang, J. & Bonami, J. R. Purification and characterization of a new reovirus from the Chinese Mitten Crab, *Eriocheir sinensis*. *J. Fish Dis.* **27**, 687–692 (2004).
- Chen, F., Suttle, C. A. & Short, S. M. Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. *Appl. Environ. Microbiol.* **62**, 2869–2874 (1996).
- Short, S. M. & Suttle, C. A. Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. *Appl. Environ. Microbiol.* **68**, 1290–1296 (2002).
- Wilson, W. H., Fuller, N. J., Joint, I. R. & Mann, N. H. Analysis of cyanophage diversity and population structure in a south-north transect of the Atlantic Ocean. *Bull. Inst. Oceanogr.* **19**, 209–216 (1999).
- Frederickson, C. M., Short, S. M. & Suttle, C. A. The physical environment affects cyanophage communities in British Columbia inlets. *Microb. Ecol.* **46**, 348–357 (2003).
- Zhong, Y., Chen, F., Wilhelm, S. W., Poorvin, L. & Hodson, R. E. Phylogenetic diversity of marine cyanophage isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene G20. *Appl. Environ. Microbiol.* **68**, 1576–1584 (2002).
- Short, C. M. & Suttle, C. A. Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. *Appl. Environ. Microbiol.* **71**, 480–486 (2005).
- Breitbart, M., Miyake, J. H. & Rohwer, F. Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol. Lett.* **236**, 249–256 (2004).
- Culley, A. I., Lang, A. S. & Suttle, C. A. High diversity of unknown picorna-like viruses in the sea. *Nature* **424**, 1054–1057 (2003).
- Breitbart, M. *et al.* Genomic analysis of uncultured marine viral communities. *Proc. Natl Acad. Sci. USA* **99**, 14250–14255 (2002).
- Breitbart, M. *et al.* Diversity and population structure of a near-shore marine-sediment viral community. *Proc. R. Soc. Lond. B* **271**, 565–574 (2004).
- Breitbart, M. & Rohwer, F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* **13**, 278–284 (2005).
- Edwards, R. A. & Rohwer, F. Viral metagenomics. *Nature Rev. Microbiol.* **3**, 504–510 (2005).
- Rohwer, F. *et al.* The complete genomic sequence of the marine phage *Roseophage* SIO1 shares homology with nonmarine phages. *Limnol. Oceanogr.* **45**, 408–418 (2000).
- Hardies, S. C., Comeau, A. M., Serwer, P. & Suttle, C. A. The complete sequence of marine bacteriophage VpV262 infecting *Vibrio parahaemolyticus* indicates that an ancestral component of a T7 viral supergroup is widespread in the marine environment. *Virology* **310**, 359–371 (2003).
- Chen, F. & Lu, J. R. Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. *Appl. Environ. Microbiol.* **68**, 2589–2594 (2002).
- Paul, J. H. *et al.* Complete genome sequence of φHSLC, a pseudotemperate marine phage of *Listonella pelagia*. *Appl. Environ. Microbiol.* **71**, 3311–3320 (2005).
- Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F. & Chisholm, S. W. Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biol.* **3**, 790–806 (2005).

63. Wilson, W. H. *et al.* Complete genome sequence and lytic phase transcription profile of a *Coccolithovirus*. *Science* **309**, 1090–1092 (2005).
64. Mann, N. H., Cook, A., Millard, A., Bailey, S. & Clokie, M. Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature* **424**, 741 (2003).
65. Millard, A., Clokie, M. R. J., Shub, D. A. & Mann, N. H. Genetic organization of the psbAD region in phages infecting marine *Synechococcus* strains. *Proc. Natl Acad. Sci. USA* **101**, 11007–11012 (2004).
66. Lindell, D. *et al.* Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl Acad. Sci. USA* **101**, 11013–11018 (2004).
67. Zeidner, G. *et al.* Potential photosynthesis gene swapping between *Prochlorococcus* and *Synechococcus* via viral intermediates. *Environ. Microbiol.* published online 23 June 2005 (doi: 10.1111/j.1462-2920.2005.00833.x).
68. Hambly, E. & Suttle, C. A. The virosphere, diversity and genetic exchange within phage communities. *Curr. Opin. Microbiol.* **8**, 444–450 (2005).
69. Heldal, M. & Bratbak, G. Production and decay of viruses in aquatic environments. *Mar. Ecol. Prog. Ser.* **72**, 205–212 (1991).
70. Suttle, C. A. & Chen, F. Mechanisms and rates of decay of marine viruses in seawater. *Appl. Environ. Microbiol.* **58**, 3721–3729 (1992).
71. Noble, R. T. & Fuhrman, J. A. Rapid virus production and removal as measured with fluorescently labeled viruses as tracers. *Appl. Environ. Microbiol.* **66**, 3790–3797 (2000).
72. Steward, G. F., Wikner, J., Cochlan, W. P., Smith, D. C. & Azam, F. Estimation of virus production in the sea: 2. Field results. *Mar. Microb. Food Webs* **6**, 79–90 (1992).
73. Wilhelm, S. W., Brigden, S. M. & Suttle, C. A. A dilution technique for the direct measurement of viral production: a comparison in stratified and tidally mixed coastal waters. *Microb. Ecol.* **43**, 168–173 (2002).
74. Evans, C., Archer, S. D., Jacquet, S. & Wilson, W. H. Direct estimates of the contribution of viral lysis and microzooplankton grazing to the decline of a *Micromonas* spp. population. *Aquat. Microb. Ecol.* **30**, 207–219 (2003).
75. Suttle, C. A. The significance of viruses to mortality in aquatic microbial communities. *Microb. Ecol.* **28**, 237–243 (1994).
76. Fuhrman, J. A. & Noble, R. T. Viruses and protists cause similar bacterial mortality in coastal seawater. *Limnol. Oceanogr.* **40**, 1236–1242 (1996).
77. Middelboe, M., Riemann, L., Steward, G. F., Hansen, V. & Nybroe, O. Virus-induced transfer of organic carbon between marine bacteria in a model community. *Aquat. Microb. Ecol.* **33**, 1–10 (2003).
78. Lawrence, J. E. & Suttle, C. A. Effect of viral infection on sinking rates of *Heterosigma akashiwo* and its implications for bloom termination. *Aquat. Microb. Ecol.* **37**, 1–7 (2004).
79. Gobler, C. J., Hutchins, D. A., Fisher, N. S., Cosper, E. M. & Sanudo-Wilhelmy, S. Release and bioavailability of C, N, P, Se, and Fe following viral lysis of a marine Chrysophyte. *Limnol. Oceanogr.* **42**, 1492–1504 (1997).
80. Poorvin, L., Rinta-Kanto, J. M., Hutchins, D. A. & Wilhelm, S. W. Viral release of iron and its bioavailability to marine plankton. *Limnol. Oceanogr.* **49**, 1734–1741 (2004).
81. Daughney, C. J. *et al.* Adsorption and precipitation of iron from seawater on a marine bacteriophage (PWH3a-P1). *Mar. Chem.* **91**, 101–115 (2004).
82. Smith, A. W. *et al.* Antisense treatment of Caliciviridae: an emerging disease agent of animals and humans. *Curr. Opin. Mol. Ther.* **4**, 177–184 (2002).
83. Philippa, J. D. W. *et al.* Antibodies to selected pathogens in free-ranging terrestrial carnivores and marine mammals in Canada. *Vet. Rec.* **155**, 135–140 (2004).
84. Li, K. S. *et al.* Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature* **430**, 209–213 (2004).
85. Delaroque, N. *et al.* The complete DNA sequence of the *Ectocarpus siliculosus* Virus EsV-1 genome. *Virology* **287**, 112–132 (2001).

Acknowledgements I have an unrepayable debt of gratitude to L. M. Proctor and J. A. Fuhrman for introducing me to the world of viruses in the ocean and how to study them, and to A. M. Chan for her critical attention to detail and scientific excellence. In addition, I have been fortunate to have been continuously educated by the many scientific colleagues that have chosen to work in my laboratory over the years, and contributed to a rich and stimulating environment. I thank R. G. Hendrix of PBI for the idea of scaling viral abundance to planetary proportions, and giving permission to use the idea in this paper. Finally, I apologize for all the excellent scientific contributions that I have been unable to include in this abbreviated format.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence should be addressed to C.A.S. (csuttle@eos.ubc.ca).

Polar ocean ecosystems in a changing world

Victor Smetacek¹ and Stephen Nicol^{2,3}

Polar organisms have adapted their seasonal cycles to the dynamic interface between ice and water. This interface ranges from the micrometre-sized brine channels within sea ice to the planetary-scale advance and retreat of sea ice. Polar marine ecosystems are particularly sensitive to climate change because small temperature differences can have large effects on the extent and thickness of sea ice. Little is known about the interactions between large, long-lived organisms and their planktonic food supply. Disentangling the effects of human exploitation of upper trophic levels from basin-wide, decade-scale climate cycles to identify long-term, global trends is a daunting challenge facing polar bio-oceanography.

Polar marine ecosystems are located at the ends of planetary gradients in daily and annual solar radiation and are ice-covered for varying lengths of the year. They harbour, or have until recently, large stocks of conspicuous animal life — birds, seals and whales — which led to the conclusion that polar ecosystems channel a greater proportion of primary production to warm-blooded predators than those at lower latitudes¹. This early view was explained by short, low-diversity food chains in polar regions, evoking images of simple systems dominated by a few key organisms.

Research conducted over the past two decades has shown that the concept of short, low-diversity polar food chains is overly simplistic. Although comparatively few species do provide the bulk of food to polar marine predators, the planktonic base of their food supply is equivalent in phylogenetic diversity to the planktonic base in temperate climate zones^{2,3}, implying that the structure and functioning of pelagic (open-water) food webs are broadly similar across all latitudes. But the key prey organisms for vertebrates vary between polar ecosystems: benthos and fish are the main prey in the north whereas crustaceans are in the south. This indicates that cold adaptation has not favoured a specific food chain. If air-breathing predators play a greater role in polar-ecosystem functioning than they do in lower latitudes, then a decrease in their stocks should have cascading effects down the food chain and lead to marked shifts in ecosystem structure. The evidence for such cascading effects is controversial, however, because of both the absence of baselines against which to assess change⁴ and our poor understanding of the carrying capacity of pelagic food webs for higher trophic levels.

Are seasonally ice-covered pelagic ecosystems fundamentally different from their counterparts in adjacent ice-free waters and how will they be affected by the retreat of sea ice in a warming world? Clearly the organisms that live in the sea ice or are dependent on it to complete their life cycles will be most severely affected, but some organisms may actually benefit from sea-ice retreat and overall productivity might actually increase⁵.

In this review we examine the possible effects of a warming world on polar ecosystems and consider only the seas and oceans directly influenced by sea ice and its melting. Because human-mediated change has influenced polar ecosystems at both ends — thinning and retreat of the ice cover and heavy exploitation of top predator populations — unrav-

elling the effects of bottom-up and top-down forcing on pelagic ecosystems is an immediate task facing polar bio-oceanographers.

Polar marine predators

It is widely assumed that marine ecosystems are controlled by bottom-up processes: the supply of nutrients by physical forcing and their conversion to organic matter by phytoplankton photosynthesis. Higher trophic levels simply harvest the levels below them. However, there is no simple relationship between the magnitude of primary production and the biomass and composition of higher trophic levels. That ecosystem interactions are more complex is demonstrated by the marked, decade-scale fluctuations in upper-level predator populations of all high-latitude systems that seem to be linked to corresponding changes in local climate and hydrography. Because the effects of climate change are difficult to disentangle from those of human exploitation, and because the sea-ice environment can directly affect both productivity and the population dynamics of birds and mammals, the mechanisms by which environmental changes are translated up the polar food web have remained elusive. An example from terrestrial ecology is instructive in this context.

The mammalian fauna of northern Eurasia and America, until the end of the last ice age, was morphologically and functionally similar to the extant megafauna of Africa and India. Herds of mammoth, deer, horses and bison, preyed upon by lions and wolves, roamed a landscape covered with a vegetation called mammoth steppe that became extinct together with the megafauna. The megafauna is likely to have conditioned the vegetation at the landscape level by promoting the growth of grasses on which they depended^{6,7}. It seems increasingly likely that human exploitation rather than climate change was responsible for the global demise of terrestrial megafauna⁸ and consequently its characteristic vegetation⁶. It seems that the wave of extinction proceeded from temperate to Arctic regions. Dwarf mammoths survived on Wrangel Island in the high Arctic and were hunted by humans till only 5,500 years ago⁹.

This insight from terrestrial ecology can assist our understanding of marine ecosystems and the effect of past and ongoing human depredation¹⁰. First, the wave of megafaunal extinction associated with human expansion in Europe will have also affected the accessible

¹Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany; ²Australian Antarctic Division, Department of the Environment and Heritage, Channel Highway, Kingston, Tasmania 7050, Australia; ³Antarctic Climate and Ecosystems Cooperative Research Centre, University of Tasmania, Hobart, Australia.

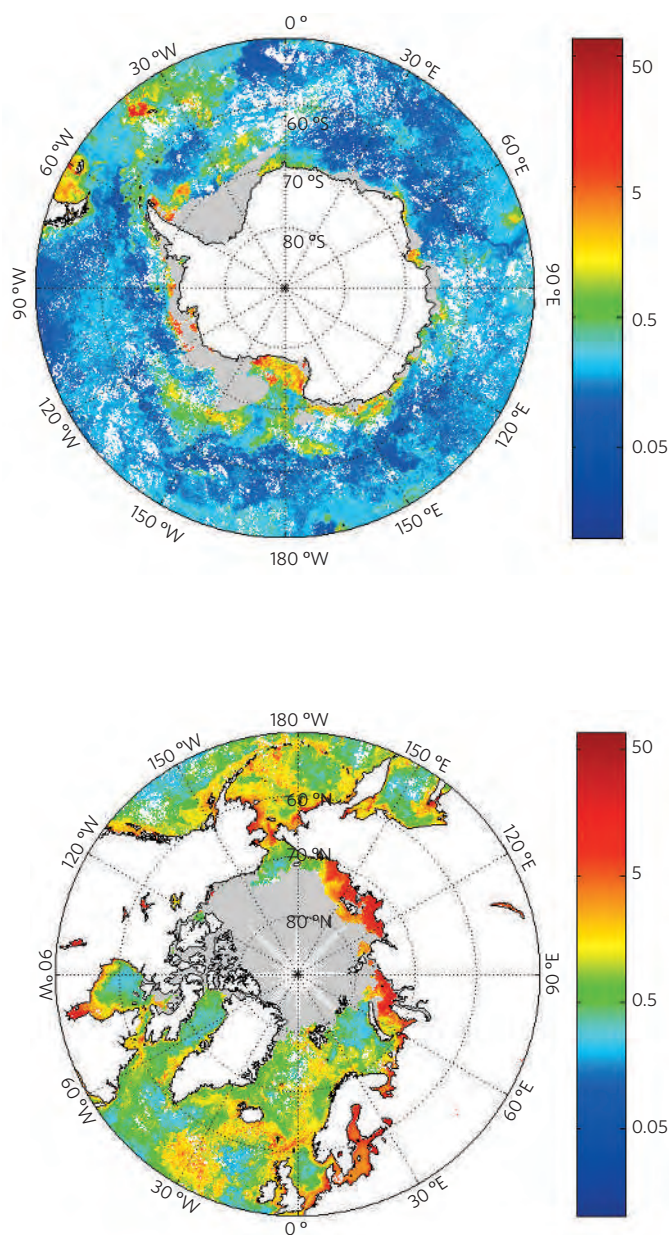


Figure 1 | Composite SeaWiFS images of the Antarctic and Arctic summer of 2002 showing surface chlorophyll *a* concentration in mg m^{-3} . Blue and green indicate low chlorophyll concentrations; reds and oranges are high concentrations. Note the highest levels in coastal areas or in areas in the lee of island groups. The high values along the Siberian shelf are due to turbidity not caused by phytoplankton. White is missing data and grey shows sea ice extent.

stocks of marine mammals and birds. Temperate populations of marine mammals were probably as large as comparable unexploited polar ecosystems¹¹. The European grey whale and the north Pacific sea cow were hunted to extinction in the seventeenth and nineteenth centuries, respectively¹¹. Second, marine megafauna, that is, mammals, birds, fish and squid, may be capable of conditioning their aquatic environment, albeit in ways that are very different from their terrestrial counterparts. Thus, terrestrial megafauna are largely herbivorous whereas marine megafauna are predators. How such feedback mechanisms might operate across trophic levels and affect pelagic ecosystem functioning is not clear. Third, it might be too late to unravel the mechanisms of top-down control on marine ecosystems, because the human onslaught on the megafauna continues to proceed unabated

from the land to the sea⁴. The polar regions may be the last refuge for the marine megafauna — the Serengeti of the sea — implying that their teeming animal life is but a reminder of what the temperate oceans were like before human exploitation. These refuges are now threatened by continued exploitation and the effects of global warming on the sea-ice environment.

The effects of the decline in sea ice on polar predators will vary with the respective degree of life-cycle adaptation to seasonality in food supply. Understanding how populations of mobile predators, from fish to whales, are geared to carrying capacities of their environment is a prerequisite if we are to achieve truly sustainable exploitation of large, long-lived organisms as well as their smaller pelagic prey. Carrying capacity will be determined by the food supply of the adult populations, which are more resilient to interannual variation than their juveniles. Relatively stable predator populations will, in turn, exert a stabilizing effect on populations of shorter-lived organisms of lower trophic levels, which tend to be more susceptible to changes in the physical environment. If predators can stabilize their population size around a long-term average food supply, then depletion of their populations through harvesting should result in a decrease in their feeding pressure and changes in food-web structure. Evidence of such a trophic cascade has been reported in a cod-dominated northern ecosystem¹², and trophic cascades may prove to be the rule rather than the exception. Investigation of top-down effects in marine ecosystems can best be carried out in polar regions because of their shorter history of exploitation and their larger extant megafaunal populations.

Polar pelagic ecosystems

The range of marine habitats represented in both polar regions is as broad as that of adjacent temperate climate zones. The north is characterized by extensive, shallow shelf seas surrounding a largely land-locked ocean, whereas the south is dominated by a land-remote, dynamic, open ocean and a very deep continental shelf. Interaction between topography, hydrography and the dynamics of the marine and terrestrial ice covers shapes these habitats. Polar marine habitats are occupied by organisms related to their counterparts in lower latitudes. Indeed, almost all major prokaryote and eukaryote lineages are represented in pelagic food webs of polar waters, indicating that they have successfully adjusted their physiological rates and life cycles to gradual cooling of these habitats while maintaining their functional role within them³. Clearly, speciation in the polar realms is driven by much the same combination of environmental conditions and organism interactions as in warmer regions. It follows that ecosystem structures are also similar, with the exception, of course, of the unique polar sea-ice habitat.

Sea ice is permeated by systems of brine channels inhabited by ice biota. This community is generally dominated by pennate diatoms, an algal group that also dominates pore waters of sandy sediments. Adaptation to confinement in narrow spaces is clearly a requirement for the sea-ice habitat, as is the ability to withstand supercooled brine in upper reaches of ice floes. Relatively few species are obligate ice inhabitants, because sea-ice is ephemeral and needs to be annually recolonized from the sea. The distances and time spans spent in water will be short in perennial ice fields and long in regions with short-lived seasonal sea-ice covers. There are accordingly large biogeographical differences in composition of sea-ice biota¹³. Biomass concentrations of microorganisms in sea ice can far exceed those attained in seawater, but their contribution to regional productivity tends to be minor because of the small relative volume of their habitat.

The effect of sea ice on pelagic systems is twofold. First, light penetration into the water column is reduced, negatively affecting underlying pelagic production. Second, in spring, microorganisms and trace elements, particularly iron, which were incorporated in the ice during formation or accumulated as dust in the snow cover, are released into the meltwater-stabilized surface layer. This process accelerates the spring bloom; however, there are great regional differences in the

amount of iron and algae released by sea ice hence the stimulating effect of meltwater on productivity. Under optimal conditions, intense, short-lived phytoplankton blooms grow in shallow surface layers in the wake of the retreating ice, seeded by iron and algal cells from the ice that tend to sink out of the surface layer following nutrient exhaustion¹³. These blooms are similar to the spring blooms of temperate shelves and divert potential food from pelagic to benthic systems¹⁴.

Despite lower temperatures and stronger seasonality, the productivities of polar ecosystems are similar to those of lower latitudes. The maximum growth rates of polar phytoplankton at *in situ* temperatures are lower than those of their temperate counterparts, but their photosynthetic efficiencies (the proportion of captured photons channelled to photosynthesis) are similar, implying that they are shade adapted¹⁵. Light intensities decrease with increasing latitude, but this dimming effect is compensated for by the increasing summer day length. Indeed, daily radiation available for photosynthesis during the polar summer is about the same as in the tropics, so primary production rates per day measured in some Arctic blooms rank among the highest values ($>5 \text{ g C m}^{-2}$) recorded anywhere¹⁵. Broad patterns of annual productivity and phytoplankton species succession are similar across the temperate to polar gradient. Differences can be attributed to seasonality of sea-ice cover and water mass properties rather than to temperature in itself. Apart from areas under perennial ice cover, polar phytoplankton growth is eventually limited during the course of the growth season, as elsewhere, by the supply of iron or macronutrients.

Although growth rates of bacteria and heterotrophic protists (protozoa), which graze on bacteria, decrease with temperature, they do not appear to play a lesser role in polar ecosystems than in lower latitudes^{16,17}. Polar macroherbivores have longer life cycles than their temperate counterparts¹⁸ but community composition and biomass vary with regional productivity rather than temperature gradients¹⁹. Relatively few pelagic herbivores enter winter dormancy, even under ice cover; instead, many maintain a level of activity during the winter commensurate with the dwindling food supply^{19,20}. Several species of copepods, amphipods and two species of Antarctic euphausiid (krill) have adapted to life in close association with sea ice, where they feed on ice algae and seek shelter from predators. It is these organisms and their predators that are being affected most by warming.

Summing up, the latitudinal transition from ice-free to ice-covered oceans is not marked by a significant shift in pelagic ecosystem structure and functioning: polar pelagic food webs are not shorter or simpler than elsewhere. However, the ecosystems encircling Antarctica do differ in some respects from those of other oceans, including the Arctic, especially with regard to the role of fish and the structure of benthic communities. Given the fundamental differences in topography, hydrography and history of glaciation it is not surprising that the Arctic and Antarctic oceans will respond differently to a warming world.

The northern polar regions

At its seasonal maximum, northern sea ice covers $13.9 \times 10^6 \text{ km}^2$, extending as far south as 44° latitude in northern Japan and covering most of the extensive continental shelves in the Pacific. Interestingly, the deep southwestern half of the Bering Sea remains ice-free, whereas the northeastern half, located over a shallow, gently sloping shelf some 700 km wide, used to be entirely covered with ice before the 1990s²¹. In the Atlantic, the winter sea-ice margin lies north of 70°N in the Barents Sea. Perennial sea ice is restricted to the Arctic Ocean where it covers $6.2 \times 10^6 \text{ km}^2$, which is equivalent to 42% of the area. Its greatest density lies north of the Canadian archipelago and Greenland. This ice field, comprising multi-year ice floes, is gradually exported south along the Greenland coast and replaced by ice forming along the exposed Siberian shelves. The Arctic multi-year ice field has thinned since the 1970s, as a result of either local melting or increased export due to a change in circulation²².

Much of the northern sea ice is formed over shallow shelves that comprise about half the total area of the Arctic Ocean²¹. Sea-ice for-

mation is a dynamic process that, in turbid coastal waters, leads to incorporation of resuspended sediment particles in the ice floes. Some of the iron contained in these sediment particles will be available to phytoplankton upon melting, and the intense Arctic and subarctic blooms that grow until nitrate is depleted in the wake of the retreating ice edge indicate that there is no lack of iron. By contrast, the ice-free, deep southwestern half of the Bering Sea is a typical high-nutrient, low-chlorophyll (HNLC) region and is clearly iron limited (Fig. 1).

Productivity under perennial ice cover in the central Arctic Ocean is constrained by light supply rather than nutrients. The ice-based food chain of the high Arctic, which leads from ice biota to copepods and amphipods, polar cod, seals to polar bears, will suffer from shrinking of the sea-ice cover even though greater areas of open water may enhance overall oceanic productivity¹⁴. However, any such increase is likely to be modest because of low nutrient concentrations, which are at the lower end of the range reported from high latitude oceans²¹. The Arctic Ocean is impoverished in nutrients because strong, haline stratification prevents admixture of deep, nutrient-rich water either by deep winter convection or upwelling. Buoyancy of the surface layer is maintained by the discharges of many rivers (of low nutrient content) and melting of sea ice advected from the coasts. How changing sea-ice dynamics in a warming world will affect the complex stratification of the Arctic Ocean is uncertain²².

The two gateways to the Arctic Ocean, the Barents Sea in the Atlantic and the Chukchi Sea in the Pacific sectors, are much more productive than the Siberian and Canadian shelf seas because of advected nutrients which are retained in the system. Water flowing through the Bering Strait originates from the deep, southwestern Bering Sea, which is rich in nutrients owing to iron limitation. Iron is supplied by admixture with shelf water and from melting ice before and during passage through the shallow (approx. 50 m) strait, rendering this region one of the most productive in the world ocean²¹. Because of its shallow depth, much of the bloom biomass settles out on the sediments where it is used by a rich benthic fauna — the feeding grounds of grey whale and walrus. Nutrient recycling between the water column and sediments is intense. The Barents Sea, in contrast, is deeper and a greater proportion of phytoplankton production is retained in the water column. The exceptionally high copepod biomass of the western Barents Sea, largely attributable to advection from the adjoining Norwegian Sea, supports, or has supported, huge stocks of planktivorous fish — capelin and herring — and their predators, particularly Atlantic cod.

The Bering and Barents Seas differ fundamentally from one another in topography and hydrography, but both maintain, or have maintained, exceptionally large stocks of fish that have been subjected to heavy exploitation²¹. In the Bering Sea, large populations of marine birds and mammals (seals, walrus and whales) feed on fish and benthos but the comparable paucity of these higher predators in the Barents Sea is probably due to heavy human exploitation reaching back to prehistoric times. Because of the warmer climate, European hunters could tide over bad hunting years with subsistence farming. This was not possible in the Bering Sea where human populations were entirely dependent on marine food, in particular walrus, and so maintained a sustainable balance with their food supply for the past millennia. Climate change has severely disrupted their lifestyle in the past decades and is the cause of serious concern today²³.

Fish stocks in both seas have fluctuated more than tenfold in the past decades partly as a result of fishing pressure but also because of basin-scale climate oscillations accompanied by shifting wind fields, storm frequency, sea-ice cover and circulation patterns^{21,24}. Productivity in the Barents Sea is higher in the periods when more warm Atlantic water is advected onto the shelf supplying nutrients and copepods. The situation appears much more complex in the Bering Sea, possibly because its enclosed relatively homogeneous topography and large predator populations leave more scope for top-down-driven oscillations than the advection-dominated regime of the Barents Sea.

Primary productivity can be estimated from satellites but assessing population sizes of fish and their predators is much more difficult, albeit a prerequisite for assessing ecosystem carrying capacity and feedback loops within the biota²³. A concerted, interdisciplinary effort to carry out comparative studies of the structure and functioning of ecosystems of subarctic seas is currently underway²⁵.

Climate models indicate that the trend in poleward retreat of Arctic sea ice is now decoupled from natural, decade-scale cycles and is due to anthropogenic forcing²⁶. The same models predict that, by the end of the century, the Arctic Ocean will be predominantly ice-free in summer. The level of productivity in an ice-free water column will depend on the extent of nutrient replenishment by vertical mixing during winter. If the mixing depth remains as shallow as under current stratification, it is feasible that productivity of the deep Arctic will decline with depletion of nitrogen and phosphorus inventories through organic particles sinking to the deep ocean.

The ice-covered Southern Ocean

The Antarctic ice cap extends into the ocean along much of the coast. Its weight presses Antarctica 200 m below the level of ice-free continents, and so its continental shelf lies considerably deeper than shelves in other oceans. Extensive shelves, partly covered by floating continental ice, occur only in the Ross and Weddell Seas. Their average depth exceeds 500 m. Land–ocean–atmosphere interaction is severely curtailed by the ice cap, and ‘normal’ mechanisms for supplying iron to the ocean through rivers and dust are absent. Nutrient input from sediments is restricted to shallow inshore environments. Antarctic sea ice, formed over deep water, is sediment-free, so Southern Ocean productivity is iron-limited, and concentrations of macronutrients are rarely exhausted.

Productive regions are restricted to the Antarctic continental margin and only extend offshore where water enriched with iron from land contact or from upwelling along shelves and continental slopes mixes with oceanic water impoverished in iron (Fig. 1). The largest of these regions stretches eastward from the tip of the Antarctic Peninsula towards South Georgia in the Southwest Atlantic sector. The Antarctic krill population is concentrated here, as are the populations of its warm-blooded predators²⁷. Where iron is supplied from contact with land-masses, dense ice-edge blooms follow the seasonal melting of the sea ice. These blooms tend to be patchy at the mesoscale (tens to hundreds of km), reflecting underlying hydrography and regional variations in ice algae distribution (Fig. 2). Further, dust settling with snow on ice floes is likely to be another significant source of iron in meltwater²⁸ that could cause similar patchiness in local phytoplankton density.

Apart from supplying iron, melting sea ice stabilizes a shallow mixed layer (< 40 m), which facilitates accumulation of bloom biomass. However, blooms also occur in deep (> 60 m) wind-mixed layers^{3,29}, and artificial blooms induced in iron fertilization experiments indicate that iron availability, rather than shallowness of the mixed layer, is the primary prerequisite for bloom formation^{30–33}. Integrated bloom biomass in deep mixed layers can rival or exceed those of shallower layers on an area by area basis although the lower concentrations can reduce grazing rates and favour larger herbivores capable of processing greater volumes. Thus a greater proportion of primary production of deep blooms is likely to be retained within the mixed layer than in shallow blooms.

The dominant algae in shallow blooms tend to be weakly silicified, fast-growing diatom species, whereas larger, grazer-resistant, heavily silicified ones, together with small flagellates, dominate deep blooms³. Because frustules of dead diatoms take longer to dissolve than their plasma, biogenic silica reaches greater depths than organic nitrogen and phosphorus. The thicker the frustules, the deeper they are likely to sink before dissolution. Dissolved silicon concentrations in the deep Southern Ocean are exceptionally high because the combination of circumpolar circulation and heavily silicified diatoms retains silicon in deep waters whereas N and P are exported to other oceans in sur-

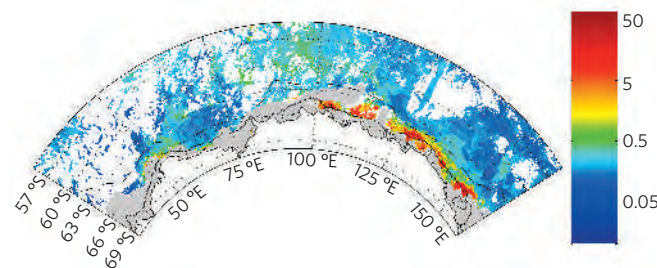


Figure 2 SeaWiFS image indicating patchiness in the ice edge bloom in January 2002 off east Antarctica. The sea ice is shown in grey.

face water downwelling north of the Polar Front³⁴. A southward retreat of winter sea ice will diminish the areal extent of dense, shallow blooms and increase that of deep blooms with ramifying effects on biogeochemical cycles and grazer populations. Changes in species composition of the phytoplankton community and the seasonality of production³⁵ can affect food web structure. Geochemical proxies indicate that productivity in the seasonal sea-ice zone (SIZ) was lower during glacial periods when winter sea-ice extended some 5° further North in the Atlantic than today and reached South Georgia^{36,37}. This cannot be attributed to unfavourable growth conditions given the higher glacial-iron deposition rates on continental ice and in the sediments³⁸. One possibility is that production in the glacial SIZ was higher but was recycled more efficiently by larger populations of grazers and predators.

The fate of bloom biomass depends on the rate of recycling of organic material including iron in the surface layer³. Heavily grazed blooms, whether deep or shallow, are likely to retain more biogenic elements in the surface layer than ungrazed blooms because the bulk of faecal material is recycled in the surface layer whereas ungrazed diatoms tend to sink out of it^{3,39}. However, mass defecation by dense swarms of krill or salps could swamp the recycling system and contribute to vertical flux⁴⁰. Faecal material of many marine mammals and birds tends to be loose and is dispersed by wave action at the surface (Fig. 3), which also contributes to the recycling pool. Adult predators convert much of their food intake into lipid reserves, retaining energy but releasing essential elements back into the system. Solar energy flowing through the food web in the course of recycled production can thus accumulate as lipids in predators. The larger the animal stock, the greater the proportion of the bloom that is retained and recycled in the surface layer, prolonging the lifetime of essential elements, particularly iron, in the productive layer — is this the marine equivalent of terrestrial ecosystem conditioning?

The major Southern Ocean macroherbivores are copepods, salps and krill, which are often spatially segregated⁴¹. Antarctic copepods have life cycles of many months to 2 years. Salps reproduce rapidly through both sexual and asexual means but are more abundant in impoverished HNLC waters. Antarctic krill (*Euphausia superba*) have a longer life cycle (5–7 years) and are thus affected by processes on wider temporal and spatial scales⁴². Because of their pivotal role in transfer of primary production to higher trophic levels, krill have received much more attention than other Antarctic herbivores. Nevertheless, there are still some fundamental gaps in our understanding of krill biology and population dynamics.

Krill links microorganisms to whales

Antarctic krill is a remarkably versatile animal that provides the food base of a range of predators from birds, seals and whales to fish and squid because it exploits both the sea-ice and open-water habitats. In general, krill distribution overlaps with that of winter sea ice, except around South Georgia, but highest densities are found in shelf-break areas with high summer chlorophyll concentrations²⁷.



Figure 3 | A minke whale off the Antarctic Peninsula dispersing nutrients in the surface layer after feeding on krill. The whale's body is covered in a thick diatom layer. (Photograph provided by J. Brokowski.)



Figure 4 | A swarm of krill at the surface. Krill can occupy all layers of the water column which makes them prey to surface, pelagic and benthic feeders. (Photograph by S. Nicol.)

The linkage between the annual abundance and recruitment success of krill and the interannual variation in the extent of sea ice⁴³ is thought to result from the greater winter feeding habitat offered by extensive sea ice and the effect of sea-ice melt on the subsequent spring bloom. The relationship, however, is not simple and there are both annual and regional exceptions^{42,44}. Although the distribution pattern of krill is relatively well established, its overall biomass around Antarctica remains uncertain. Calculations based on potential consumption of primary production by krill suggest an upper limit of 4.4 billion tonnes, and estimates of predator consumption suggest a requirement for 150–300 million tonnes⁴⁵. Extrapolations from acoustic measurements of krill abundance indicate a krill biomass of between 60 and 155 million tonnes, whereas scientific net surveys generally produce values an order of magnitude lower⁴⁶. Analyses of all available scientific net data have indicated a potential 80% decline in krill density in the southwest Atlantic between 1976 and 2004 (ref. 27); however, this trend is difficult to verify. The only acoustic survey time series (from 1981 to present in the Elephant Island region in the SW Atlantic⁴⁴) exhibits rather different trends from net-based surveys from the same region. However, suggestions of negative trends in krill distribution

and abundance should be a cause for concern and warrant further investigation, which will have to overcome problems associated with net avoidance, uncertainty in calibration of acoustic methods, patchy distribution and regional and interannual variability in occurrence in a vast range⁴⁶.

Regional changes in krill abundance have been associated with interannual changes in the extent of sea-ice cover in areas crucial to krill recruitment²⁷, but longer-term linkages are more elusive. There is conflicting evidence for change in overall sea-ice extent, although regional changes over the past 30 years are not disputed⁴⁷. Satellite data, available from 1973, reveal considerable interannual and regional variation in sea-ice extent and concentration. There has been no consistent circumpolar trend in the extent of winter sea ice since 1973, although there may have been some changes in ice concentration⁴⁷. Along the Antarctic Peninsula, however, the extent of winter sea ice has decreased markedly in recent years: the climate in this region has been warming faster than any other part of the planet³⁵. Sea-ice proxy studies examining the pre-satellite era have suggested that a significant decrease in sea-ice extent occurred during the 1950s to 1970s^{48–50}. The likelihood of a major (~20%) decline in sea ice in the middle of the twentieth century cannot be discounted, and such a significant change in the Southern Ocean would have had major effects on the lower trophic levels, with associated ramifications for the upper level predators of the region — but again, long-term data are scarce.

Estimates of population sizes of air-breathing predators are also fraught with uncertainty: few long-term data exist for any species, and there are also severe difficulties in estimating either their local or global abundance. The krill predators expected to show some recent changes are the baleen whales, seals (largely crabeater and fur) and penguins. Decreases in baleen whale numbers have been inferred from the whaling industry in the past century⁵¹, but the process of recovery is much more difficult to detect and measure accurately. Similarly, crabeater seals were expected to take up the 'krill surplus' made available by the harvesting of baleen whales, but estimates of abundance before whaling are unreliable, as are the data to support an increase in numbers after whaling. A concerted effort to provide an up-to-date, circumpolar abundance estimate is currently underway⁵². Fur seals have been reasonably well monitored and impressive increases in abundance have been recorded at some locations, although this recovery is taking place some 150 years after the end of their exploitation⁵³. Penguins, where monitored, show a variety of both long-term and short-term population changes, which seem as much related to the presence of sea ice as to the fluctuations or changes in their food supply^{54,55}. It is difficult to distil a coherent trend for the entire suite of krill predators that might be unequivocally associated with a global or regional decline of krill, although localized studies have shown marked trends for individual species⁵⁷. This is not to say that ecosystem shifts have not taken place; rather, that these may be subtle and complex and will involve species-specific responses to their prey, their competitors and the changing environment.

Concluding remarks

Despite climate change, intense present and past resource exploitation and the unique sea-ice ecosystems, the polar regions still offer the possibility to carry out end-to-end ecological experiments in relatively undisturbed environments. Future changes in polar marine ecosystems will depend as much on global climate change as on our ability to regulate exploitation pressure at sustainable levels, and such regulation will require greater ecological certainty^{56,57}. Some physical changes are already upon us and are altering patterns of human behaviour in the northern polar regions. The livelihoods of the indigenous populations subsisting on subarctic and polar marine mammals have been severely jeopardized because their harvesting techniques tend to be tightly linked to seasonality of sea ice²³. A year-round high Antarctic krill fishing season has become a reality because of declining sea ice in the

Box 1 | The life of krill

Euphausiids are a major component of most shelf and open ocean food webs and can occupy a number of trophic levels. They are considerably larger and more active than copepods, and their size and swarming habit makes them an important link between planktonic organisms and higher trophic levels⁴³. Antarctic krill are among the larger euphausiids (up to 6 cm) and can live for 11 years (Fig. 4). Their diet is also highly varied and their complex feeding appendages enable them to collect a wide size-range of food from larger plankton (> 6 µm), particularly diatoms, to copepods. They also feed on ice biota by sweeping the underside of floes⁴⁴. Ingested food is crushed internally in a gizzard lined with teeth that shred even strong-shelled diatoms. Large, dense krill swarms can contain some 10¹² individuals, each animal consuming up to 25% of its body carbon per day, making them globally significant consumers, which can leave little in the water column after their passage⁴³. In common with shrimp, which they resemble, krill escape from predators by flips of their muscular tail, and krill swarms react to predators in much the same way as fish schools. Adult krill occur in swarms at concentrations of 30,000 animals m⁻³ and can contain up to a million tonnes of biomass (~10¹² individuals)⁴⁵. This has made them the target for the Southern Ocean's largest fishery.

Krill eggs are laid off the shelf break and sink to 700–1,000 m before hatching. The larvae undergo metamorphosis during ascent and congregate under the sea ice during winter where they feed on ice biota in this protected environment. Unlike adults, which can survive without food for many months, larval krill rapidly starve to death, so an adequate winter food supply is crucial for successful recruitment⁴⁶. High recruitment rates in the southwestern Atlantic Sector have been observed in years with extensive winter ice cover, indicating the environmental dependence of early larval stages on ice biota and dense blooms²⁷. Adult krill (2 or more years old) spend the winter mainly under the sea ice where they can feed on ice biota, feed omnivorously or starve. Because krill are long-lived animals, their population processes are attuned to cycles in the marine environment beyond the purely seasonal⁴⁷. Consequently, iron fertilization experiments aimed at studying the interactions of higher trophic levels with their herbivorous prey, and their food, would have to be at much grander scales than those conducted so far.

southwestern Atlantic⁵⁸. Hunting bans have led to recovery of some mammal stocks in both polar regions, but sadly the population of Antarctic blue whales, the largest megafauna the world has ever seen, has not yet shown significant signs of recovery³¹. Studies in polar waters tend to have a narrow research focus: biogeochemistry and climate, or biodiversity and conservation, but to understand significant changes occurring at the poles, these approaches will have to be coordinated and combined. The polar regions are changing, in some places rapidly, and where these changes are a result of resource exploitation, the rate of change can be modified through management action. Improving our ability to understand ecological processes can, hopefully, ameliorate the combined effects of climate change and the seemingly relentless human tide of marine resource exploitation. ■

- Hart, T. J. Phytoplankton periodicity in Antarctic surface waters. *Discovery Rep.* **21**, 261–356 (1942).
- Lopez-Garcia, P., Rodriguez-Valera, F., Pedros-Alio, C. & Moreira, D. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**, 603–607 (2001).
- Smetacek, V., Assmy, P. & Henjes, J. The role of grazing in structuring Southern Ocean pelagic ecosystems and biogeochemical cycles. *Antarct. Sci.* **16**, 541–558 (2004).
- Pauly, D. & Maclean, J. *In a Perfect Ocean* (Island Press, Washington, 2003).
- Arrigo, K. R. & Thomas, D. N. Large-scale importance of sea ice biology in the Southern ocean. *Antarct. Sci.* **16**, 471–486 (2004).
- Zimov, S. A. *et al.* Steppe-tundra transition: A herbivore driven biome shift at the end of the Pleistocene. *Am. Nat.* **146**, 765–794 (1995).
- Knapp, A. K. *et al.* The keystone role of bison in North American tallgrass prairie. *Bioscience* **49**, 39–50 (1999).
- Brook, B. W. & Bowman, D. M. J. S. The uncertain blitzkrieg of Pleistocene megafauna. *J. Biogeogr.* **31**, 517–523 (2004).
- Orlova, L. A., Kuzmin, Y. V. & Dementiev, V. N. A review of the evidence for extinction chronologies for five species of Upper Pleistocene megafauna in Siberia. *Radiocarbon* **46**, 301–314 (2004).
- Pauly, D., Watson, R. & Alder, J. Global trends in world fisheries: impacts on marine ecosystems and food security. *Phil. Trans. R. Soc. Lond. B* **360**, 5–12 (2005).
- Jackson, J. B. C. *et al.* Historical overfishing and the recent collapse of coastal ecosystems. *Science* **293**, 629–638 (2001).
- Frank, K. T., Petrie, B., Choi, J. S. & Leggett, W. C. Trophic cascades in a formerly cod-dominated ecosystem. *Science* **308**, 1621–1623 (2005).
- Brierley, A. S. & Thomas, D. N. Ecology of Southern Ocean pack ice. *Adv. Mar. Biol.* **43**, 171–276 (2002).
- Smetacek, V. Role of sinking in diatom life-history cycles: ecological, evolutionary and geological significance. *Mar. Biol.* **84**, 239–251 (1985).
- Sakshaug, E. In *The Arctic Carbon Cycle* (eds Stein, R. & MacDonald, R.) 58–81 (Springer, Heidelberg, 2003).
- Knox, G. A. *The Biology of the Southern Ocean* (Cambridge Univ. Press, Cambridge, UK, 1994).
- Sherr, B. F. & Sherr, E. B. Community respiration/productivity and bacterial activity in the upper water column of the central Arctic Ocean. *Deep-Sea Res.* **50**, 529–542 (2003).
- Huntley, M. E., Lopez, M. D. G. & Karl, D. M. Top predators in the Southern Ocean as a major leak in the biological carbon pump. *Science* **253**, 64–66 (1991).
- Atkinson, A. Life cycle strategies of epipelagic copepods in the Southern Ocean. *J. Mar. Syst.* **15**, 280–311 (1998).
- Ashjian, C. J., Campbell, R. G., Welch, H. E., Buter, M. & Keuren, D. V. Annual cycle in abundance, distribution and size in relation to hydrography of important copepod species in the western Arctic Ocean. *Deep-Sea Res.* **50**, 1235–1261 (2003).
- Sakshaug, E. & Walsh, J. In *The Arctic: Environment, People, Policy* (eds Nuttall, M. & Callaghan, V.) 163–196 (Harwood Acad., Amsterdam, 2000).
- Macdonald, R. W., Sakshaug, E. & Stein, R. In *The Arctic Carbon Cycle* (eds Stein, R. & MacDonald, R.) 7–22 (Springer, Heidelberg, 2003).
- Ray, G. C. & McCormick-Ray, J. *Coastal-Marine Conservation: Science and Policy* (Blackwell Science, Oxford, 2004).
- Hunt, G. L. & Stabeno, P. J. Climate change and control of energy flow in the southwestern Bering Sea. *Prog. Oceanogr.* **55**, 5–22 (2002).
- Hunt, G. L. & Drinkwater, K. F. Ecosystem studies of sub-arctic seas (ESSAS). Science Plan. GLOBEC Report No. 19, vii at <http://globec.org> (2005).
- Johannessen, O. M. *et al.* Arctic climate change: observed and modelled temperature and sea-ice variability. *Tellus A* **56**, 328–341 (2004).
- Atkinson, A., Siegel, V., Pakhomov, E. & Rothery, P. Long-term decline in krill stock and increase in salps within the Southern Ocean. *Nature* **432**, 100–103 (2004).
- Sedwick, P. N. & DiTullio, G. R. Regulation of algal blooms in Antarctic shelf waters by the release of iron from melting sea ice. *Geophys. Res. Lett.* **24**, 2515–2518 (1997).
- Turner, D. & Owens, N. J. P. A biogeochemical study in the Bellingshausen Sea: Overview of the STERNA 1992 expedition. *Deep-Sea Res.* **42**, 907–932 (1995).
- Boyd, P. W. *et al.* A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. *Nature* **407**, 695–701 (2000).
- Gervais, R., Riebesell, U. & Gorbunov, M. Y. Changes in primary productivity and chlorophyll *a* in response to iron fertilization in the Southern Polar Frontal Zone. *Limnol. Oceanogr.* **47**, 1324–1335 (2002).
- Coale, K. H. *et al.* Southern Ocean iron enrichment experiment: carbon cycling in high and low-Si waters. *Science* **304**, 408–414 (2004).
- Smetacek, V., Bathmann, U. & Helmke, E. The Expeditions ANTARKTIS XXI/3–4–5 of the research vessel 'Polarstern' in 2004. *Ber. Polarforsch. Meeresforsch.* **500**, 3–134 (2005).
- Sarmiento, J. L., Gruber, N., Brzezinski, M. A. & Dunne, J. P. High-latitude controls of thermocline nutrients and low latitude biological productivity. *Nature* **427**, 56–60 (2004).
- Moline, M. A., Claustre, H., Frazer, T. K., Schofield, O. & Vernet, M. Alteration of the food web along the Antarctic Peninsula in response to a regional warming trend. *Glob. Change Biol.* **10**, 1973–1980 (2004).
- Gersonde, R. *et al.* Last Glacial sea-surface temperatures and sea-ice extent in the Southern Ocean (Atlantic-Indian sector) — a multiproxy approach. *Paleoceanography* **18**, 1061, doi:10.1029/2002PA000809 (2003).
- Anderson, R. F., Chase, Z., Fleisher, M. Q. & Sachs, J. The Southern Ocean's biological pump during the Last Glacial Maximum. *Deep-Sea Res.* **49**, 1909–1938 (2002).
- Mahowald, N. *et al.* Dust sources and deposition during the last glacial maximum and current climate: a comparison of model results with paleodata from ice cores and marine sediments. *J. Geophys. Res.* **104**, 15895–15916 (1999).
- Gonzalez, H. E. The distribution and abundance of krill faecal material and oval pellets in the Scotia and Weddell Seas (Antarctica) and their role in vertical flux. *Polar Biol.* **12**, 81–91 (1992).
- von Bodungen, B., Fischer, G., Nöthig, E.-M. & Wefer, G. Sedimentation of krill faeces during spring development of phytoplankton in Bransfield Strait, Antarctica. *Mitt. Geol.-Palaeontol. Inst. Univ. Hamburg: SCOPE UNEP* **62**, 243–257 (1987).
- Voronina, N. M. Comparative abundance and distribution of major filter feeders in the Antarctic pelagic zone. *J. Mar. Syst.* **17**, 375–390 (1998).
- Constable, A. J., Nicol, S. & Strutton, P. G. Southern Ocean productivity in relation to spatial and temporal variation in the physical environment. *J. Geophys. Res.* **108** (C4), doi:10.1029/2001JC001270 (2003).
- Siegel, V. & Loeb, V. Recruitment of Antarctic krill *Euphausia superba* and possible causes for its variability. *Mar. Ecol.-Prog. Ser.* **123**, 45–56 (1995).
- Hewitt, R. P., Demer, D. A. & Emery, J. H. An 8-year cycle in krill biomass density inferred from acoustic surveys conducted in the vicinity of the South Shetland Islands during the Austral summers of 1991/92 through 2001/2002. *Aquat. Living Resour.* **16**, 205–213 (2003).
- Pridde, J., Boyd, I. L., Whitehouse, M. J., Murphy, E. J. & Croxall, J. P. Estimates of Southern Ocean primary production — constraints from predator demand and nutrient drawdown. *J. Mar. Syst.* **17**, 275–288 (1998).
- Nicol, S., Constable, A. & Pauly, T. Estimates of circumpolar Antarctic krill abundance based on recent acoustic density measurements. *CCAMLR Sci.* **7**, 87–99 (2000).
- Parkinson, C. L. Southern Ocean sea ice and its wider linkages: insights revealed from models and observations. *Antarct. Sci.* **16**, 387–400 (2004).
- de la Mare, W. K. Abrupt mid-twentieth-century decline in Antarctic sea ice extent from whaling records. *Nature* **389**, 57–60 (1997).
- Murphy, E. J., Clarke, A., Symon, C. & Pridde, J. Temporal variation in Antarctic sea-ice: analysis of a long term fast-ice record from the South Orkney Islands. *Deep-Sea Res.* **42**, 1–18 (1995).

50. Curran, M. A. J., van Ommen, T. D., Morgan, V. I., Phillips, K. L. & Palmer, A. S. Ice core evidence for sea ice decline since the 1950s. *Science* **302**, 1203–1206 (2003).
 51. Baker, C. S. & Clapham, P. J. Modelling the past and future of whales and whaling. *Trends Ecol. Evol.* **19**, 365–371 (2004).
 52. Southwell, C., de la Mare, B., Borchers, D. & Burt, L. Shipboard line transect surveys of crabeater seal abundance in the pack ice off eastern Antarctica: evaluation of assumptions and recommendations for analysis. *Mar. Mammal Sci.* **20**, 602–620 (2004).
 53. Reid, K. & Croxall, J. Environmental response of upper trophic-level predators reveals a system change in an Antarctic marine ecosystem. *Phil. Trans. R. Soc. Lond. B* **286**, 377–384 (2001).
 54. Fraser, W. R., Trivelpiece, W. Z., Ainley, D. G. & Trivelpiece, S. G. Increases in Antarctic penguin populations: reduced competition with whales or a loss of sea ice due to environmental warming? *Polar Biol.* **11**, 525–531 (1992).
 55. Croxall, J. P., Trathan, P. N. & Murphy, E. J. Environmental change and Antarctic seabird populations. *Science* **297**, 1510–1514 (2002).
 56. Clarke, A. & Harris, C. M. Polar marine ecosystems: major threats and future change. *Environ. Conserv.* **30**, 1–25 (2003).
 57. Croxall, J. P. & Nicol, S. Management of Southern Ocean resources as a model for global sustainability. *Antarct. Sci.* **16**, 569–584 (2004).
 58. Jones, C. D. & Ramm, D. C. The commercial harvest of krill in the southwest Atlantic before and after the CCAMLR 2000 Survey. *Deep-Sea Res. II* **51**, 1421–1434 (2004).
 59. Nicol, S. Living krill, zooplankton and experimental investigations. *Mar. Freshwat. Behav. Phys.* **36**, 191–205 (2003).
 60. Marschall, H.-P. The overwintering strategy of Antarctic krill under the pack-ice of the Weddell Sea. *Polar Biol.* **9**, 129–135 (1988).
 61. Hamner, W. M. & Hamner, P. P. Behaviour of Antarctic krill (*Euphausia superba*): schooling, foraging, and antipredatory behaviour. *Can. J. Fish. Aquat. Sci.* **57** (Suppl. S3), 192–202 (2000).
 62. Ross, R. M. Energetic cost to develop to the first feeding stage of *Euphausia superba* Dana and the effect of delays in food availability. *J. Exp. Mar. Biol. Ecol.* **133**, 103–127 (1989).
- Acknowledgements** We are grateful to A. Atkinson and G. Carleton Ray for useful discussions and B. Raymond and J. Schwarz for producing figs 1 and 2. We thank the SeaWiFS Project and the Distributed Active Archive Center at the Goddard Space Flight Center for the production and distribution of these data, respectively. These activities are sponsored by NASA's Mission to Planet Earth Program. We thank Captain J. Borkowski III of RV *Nathaniel Palmer* for permission to use fig. 3.
- Author Information** Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to V. S. (vsmetacek@awi-bremerhaven.de).

Psoriasis-like skin disease and arthritis caused by inducible epidermal deletion of Jun proteins

Rainer Zenz¹, Robert Eferl¹, Lukas Kenner¹, Lore Florin², Lars Hummerich³, Denis Mehic^{4,5}, Harald Scheuch¹, Peter Angel², Erwin Tschachler^{4,5} & Erwin F. Wagner¹

Psoriasis is a frequent, inflammatory disease of skin and joints with considerable morbidity. Here we report that in psoriatic lesions, epidermal keratinocytes have decreased expression of JunB, a gene localized in the psoriasis susceptibility region *PSORS6*. Likewise, inducible epidermal deletion of *JunB* and its functional companion *c-Jun* in adult mice leads (within two weeks) to a phenotype resembling the histological and molecular hallmarks of psoriasis, including arthritic lesions. In contrast to the skin phenotype, the development of arthritic lesions requires T and B cells and signalling through tumour necrosis factor receptor 1 (TNFR1). Prior to the disease onset, two chemotactic proteins (*S100A8* and *S100A9*) previously mapped to the psoriasis susceptibility region *PSORS4*, are strongly induced in mutant keratinocytes *in vivo* and *in vitro*. We propose that the abrogation of JunB/activator protein 1 (AP-1) in keratinocytes triggers chemokine/cytokine expression, which recruits neutrophils and macrophages to the epidermis thereby contributing to the phenotypic changes observed in psoriasis. Thus, these data support the hypothesis that epidermal alterations are sufficient to initiate both skin lesions and arthritis in psoriasis.

Psoriasis is a chronic disease of unsolved pathogenesis affecting skin and joints in 1–3% of the general population. It is characterized by inflamed, scaly and frequently disfiguring skin lesions and oligo-arthritis of the small joints of the hands and feet¹. The skin lesions show hyperproliferation and altered differentiation of epidermal keratinocytes, marked infiltrates of T cells and neutrophils, and a distinct increase of skin capillaries¹. Although at least six different psoriasis susceptibility loci, designated *PSORS1–PSORS6*, have been mapped by using genome-wide scans, the cause of psoriasis remains unknown².

JunB expression is downregulated in psoriasis

Human *JunB* (19p13.2), a component of the AP-1 transcription factor is localized in the *PSORS6* locus (19p13) and is known to regulate cell proliferation, differentiation, stress responses and cytokine expression in various organs^{3,4}. To test a possible function of JunB/AP-1 in the aetiology of psoriasis, we analysed expression of JunB and *c-Jun* in unaffected and affected areas of biopsy samples ($n = 8$) from psoriatic patients. As recently reported for healthy human skin⁵, JunB was found ubiquitously expressed in all layers of the epidermis of unaffected skin of psoriatic patients with highest levels in the basal and spinous layers (Fig. 1a). Notably, the expression of JunB was greatly reduced in lesional skin of severe psoriasis ($n = 6$) and intermediately expressed in mild psoriasis ($n = 2$; data not shown), suggesting a possible role of JunB in the development of the disease. In contrast, *c-Jun*, which is a proposed antagonist of JunB, was (with the exception of the granular layer) only weakly expressed in normal human epidermis, but quite prominent in psoriatic skin⁵ (Fig. 1a). Quantification of JunB (Fig. 1b) and *c-Jun* expression (Fig. 1c) in different layers of the epidermis showed

a strong reduction of JunB/*c-Jun* ratios in psoriatic skin, most prominently in the basal layer (Fig. 1d).

Inducible epidermal deletion of Jun proteins in adult mice

To study the consequence of downregulating AP-1 expression in the epidermis of adult mice, we designed inducible, conditional, single- and double-knockout mice for *JunB* and *c-Jun*. Mice carrying LoxP-site-containing (floxed) *JunB* and *c-Jun* alleles (Fig. 2a) were crossed to K5-Cre-ER^T transgenic mice⁶, in which tamoxifen efficiently induced Cre-mediated recombination in basal layer keratinocytes. At 8 weeks of age, the single- and double-mutant mice and their littermate controls were treated with tamoxifen (Fig. 2b; 5 d, 1 mg d⁻¹) and monitored for 18 days. Single epidermis-specific, inducible, knockout mice of *JunB* or *c-Jun* did not show any signs of a skin phenotype up to 2 months after deletion (see Supplementary Fig. 1). However, in *JunB/c-Jun* double-mutant mice, alterations to the hairless skin (that closely resembled the skin lesions of patients with psoriasis) appeared 8 to 10 days after tamoxifen induction. After 18 days of tamoxifen treatment, 100% of the double-mutant mice showed a strong phenotype with scaly plaques affecting primarily ears, paws and tail, and less frequently the hairy back skin (Fig. 2d–m). Histology of affected skin from mutant mice showed the hallmarks of psoriasis (Fig. 3a), such as a strongly thickened epidermis with prominent rete ridges, thickened keratinized upper layers (hyperkeratosis) with parakeratosis (nucleated keratinocytes in the cornified layer) and increased subepidermal vascularization. Intra-epidermal T cells, epidermal micro-abscesses and the typical inflammatory-cell infiltrate consisting of neutrophils were seen, along with increased numbers of macrophages in the dermis. Arthritic lesions strongly reminiscent of psoriatic arthritis (seen in 5–40% of psoriasis

¹Research Institute of Molecular Pathology, Dr. Bohr-Gasse 7, A-1030 Vienna, Austria. ²Division of Signal Transduction and Growth Control and ³Division of Molecular Genetics, Deutsches Krebsforschungszentrum, D-69120 Heidelberg, Germany. ⁴Department of Dermatology, Medical University of Vienna, A-1090 Vienna, Austria. ⁵Centre de Recherches et d'Investigations Épidermiques et Sensorielles, F-92511 Neuilly, France.

patients) were observed with 100% penetrance. Inflammatory infiltrates were present in the joint regions along with massive bone destruction and periostitis (Fig. 3b, c).

Molecular characterization of psoriatic phenotype

The floxed *JunB* and *c-Jun* alleles were significantly deleted in the epidermis (*JunB*^{Δep}*c-Jun*^{Δep}; data not shown) and RNase protection assay for AP-1 family members demonstrated a strong reduction of *c-Jun* and *JunB* messenger RNA in the epidermis (Fig. 2c). The residual *c-Jun* and *JunB* mRNA in the epidermis (Fig. 2c) probably originates from the presence of inflammatory cells, but also from Langerhans cells and other epidermal dendritic cells (Fig. 3). A large number of cytokines, chemokines and transcription factors have been proposed to contribute to the pathogenesis of psoriasis⁷. Therefore, the expression of several cytokines and chemokines was analysed by an RNase protection assay (Fig. 4a). Interleukin 1α (IL-1α), IL-1β, interferon-γ (IFN-γ) and TNF-α, shown to precede other cytokines and chemokines in psoriatic lesions⁸, were strongly upregulated in the diseased epidermis; whereas expression of IL-18, normally induced in differentiating keratinocytes⁹, was found to be reduced (Fig. 4a). Macrophage inflammatory protein-2 (MIP-2; IL-8 in humans), a strong chemoattractant for neutrophils and T cells mainly produced by keratinocytes on stimulation by inflammatory cytokines⁷, was also strongly expressed. In addition to MIP-2, four other chemokines were found highly expressed in mutant epidermis (Fig. 4a), MIP-1α, MIP-1β, IP-10 and monocyte chemotactic protein-1 (MCP-1), which provide a strong T-cell chemotactic effect⁷. As described for psoriatic plaques¹⁰, upregulation of IL-12p40 was detected (Fig. 4a). In addition, reduced expression of transforming growth factor-β2 (TGF-β2), a prominent inhibitor of keratinocyte proliferation¹¹, was seen in diseased epidermis, whereas TGF-β1 was unchanged (Fig. 4a). These results lend further support to the

hypothesis that the presented mouse model largely mimics human psoriasis.

For further molecular characterization of this phenotype, global gene expression profiling was performed on two different microarrays, comprising the murine 20k ArrayTAG (LION Bioscience; ref. 12) and the 22.4k NIA cDNA collection from the National Institute of Ageing, combining the 15k NIA clone set^{13,14} extended by 7.4k additional sequences¹⁵ (data not shown). Analyses of genes that were deregulated at least twofold confirmed downregulation of *JunB* and *c-Jun* in diseased epidermis of mutant mice. As reported for human psoriatic keratinocytes, the chemotactic proteins S100A8 and S100A9¹⁶ were upregulated at the RNA level (4.3-fold and 8.5-fold, respectively) and increased expression of S100A8 was detected by immunohistochemistry (Fig. 4b). Furthermore, epidermal fatty acid binding protein (Fabp5; 4.4-fold increase), secretory leukocyte protease inhibitor (Slpi; 3.3-fold increase), calmodulin (Calm1; 3.2-fold increase) and Gro1 (2.7-fold increase) were upregulated in diseased epidermis (data not shown). Similar to psoriasis, keratin 15 (ref. 17; 12.1-fold decrease) and caveolin¹⁸ (2.9-fold decrease) were found significantly reduced in mutant epidermis (data not shown). TIMP3 was more than sixfold downregulated in diseased epidermis of mutant mice (data not shown). It has been shown that TIMP3, besides inhibiting collagenase-1, stromelysin-1 and gelatinase A and B¹⁹ also strongly inhibits TACE²⁰ (TNF-α-converting enzyme) thereby controlling the release of TNFα.

Induction of chemotactic S100 proteins after Jun deletion

To identify the initiating molecular events in the aetiology of psoriasis, 8-week-old double-mutant mice and their littermates were treated with tamoxifen (3 d, 1 mg d⁻¹) and analysed 3 days after the first tamoxifen treatment. At this early time point, none of the mutant mice showed macroscopic or microscopic skin alterations

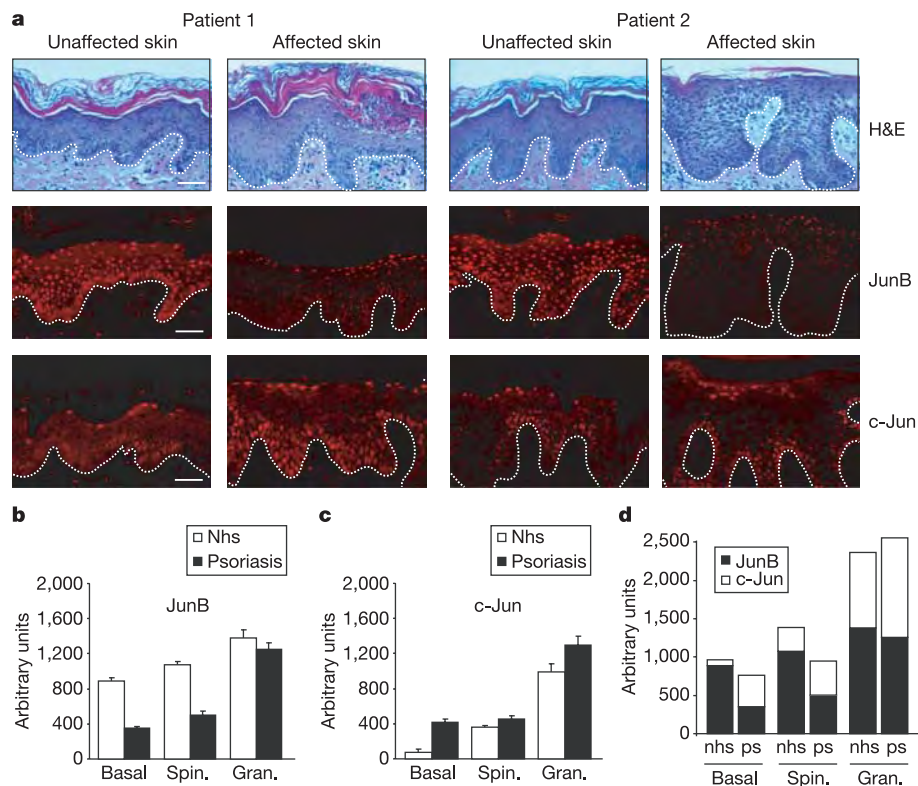


Figure 1 | Reduced JunB expression in human psoriatic skin. **a**, H&E staining and immunohistochemistry for JunB and c-Jun of unaffected and affected areas of two representative human biopsies from psoriatic patients ($n = 6$, $\times 40$). Immunohistochemistry for JunB reveals reduced expression of JunB in the epidermis of psoriatic plaques compared to unaffected skin areas of psoriatic patients. c-Jun is only weakly expressed in unaffected

epidermis, but is quite prominent in affected areas of psoriatic skin. **b–d**, Quantification of JunB (**b**) and c-Jun expression (**c**) in different layers of the epidermis, which shows a strong reduction of JunB/c-Jun ratios in psoriatic skin (**d**). Abbreviations are as follows: nhs, normal human skin; ps, psoriatic skin; basal, basal layer; Spin., spinous layer; Gran., stratum granulosum. Error bars represent the standard deviation.

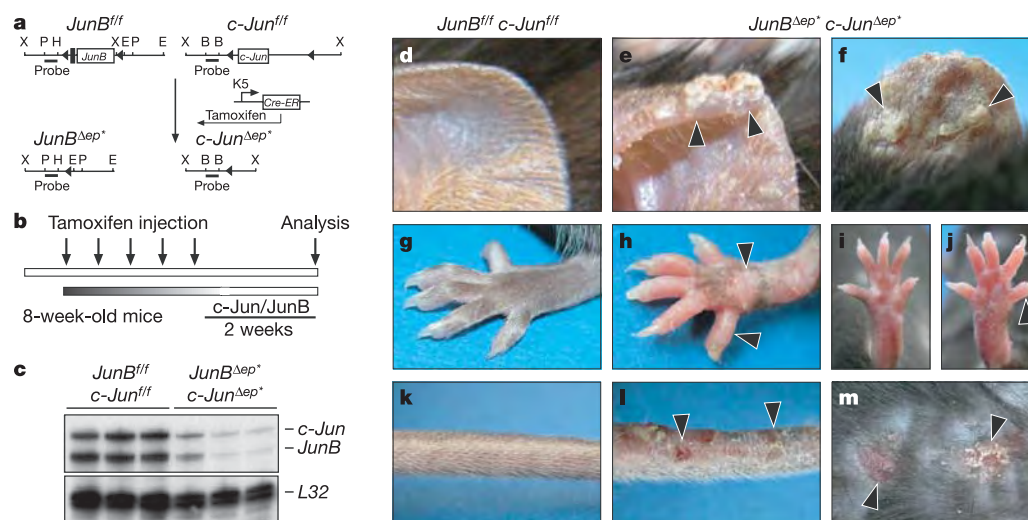


Figure 2 | Inducible deletion of *JunB* and *c-Jun* in the epidermis of adult mice. Mice carrying floxed alleles for the *JunB* and/or *c-Jun* locus were used to delete one or both genes in the epidermis by inducible Cre-recombinase activity. **a**, Schematic representation of the floxed *JunB* and *c-Jun* loci before and after tamoxifen-induced activation of the Cre-ER-fusion protein, which is under the control of the keratin 5 promoter. **b**, Eight-week-old mice were injected for five consecutive days with tamoxifen (1 mg day⁻¹, intraperitoneal). Two weeks after the last injection the double-mutant mice showed a strong skin phenotype reminiscent of psoriasis. At this time-point

samples were taken for further analyses. **c**, RNase protection assay for *c-Jun*, *JunB* and *L32* genes in three individual animals of each genotype showing reduced expression of *c-Jun* and *JunB* after inducible deletion. **d–m**, Macroscopic views of ear (**d–f**), paw (**g–j**), tail (**k, l**) and back skin (**m**) from adult mice 2 weeks after inducible deletion of *JunB* and *c-Jun* in the epidermis. Scale-like skin irritations (marked by arrows) were observed on the front and back side of the ear (**e, f**), paw (**h, j**), tail (**l**) and shaved back skin (**m**) of adult mutant mice.

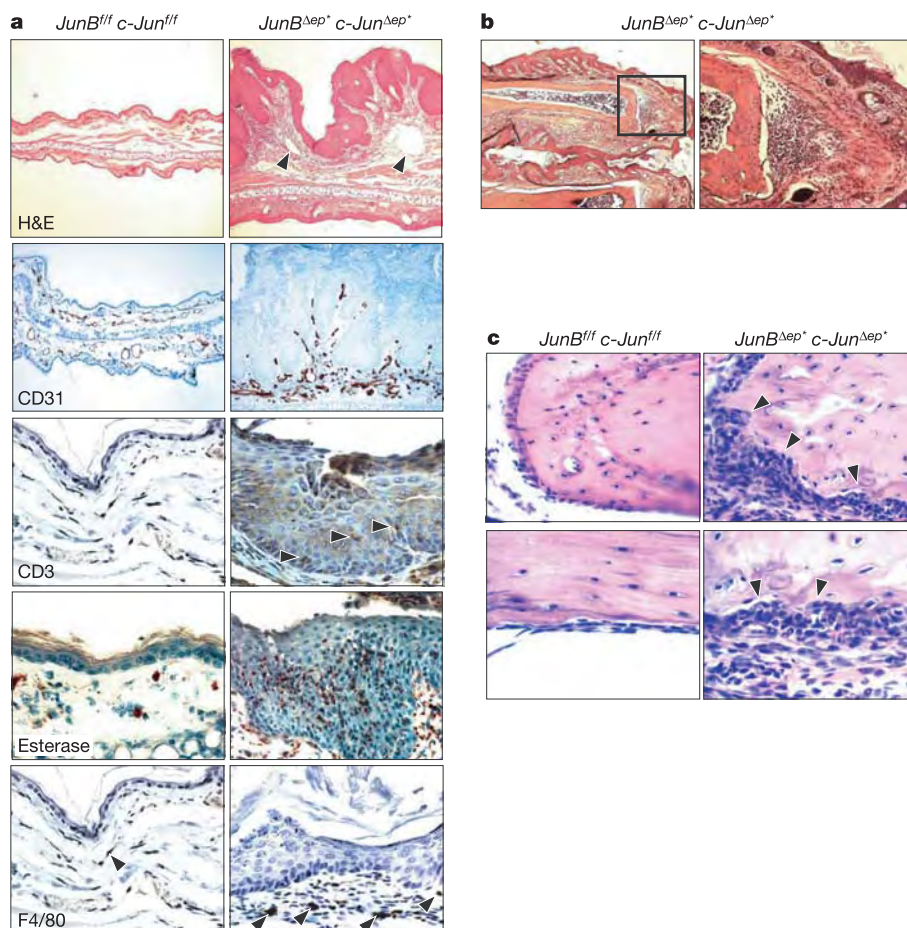


Figure 3 | Mice lacking *JunB* and *c-Jun* in the epidermis exhibit the hallmarks of psoriasis.

a, Histology of mouse ears reflecting the hallmarks of psoriasis (right panels). H&E staining shows: abnormally thickened epidermis, parakeratosis (nucleated keratinocytes in the cornified layer), thickened keratinized upper layer (hyperkeratosis), fingerlike epidermal projections into the dermis (rete ridges) and enlarged blood vessels (arrows in H&E, CD31 staining). Epidermal micro-abscesses and the typical inflammatory cell infiltrate are seen: intraepidermal T-cells (CD3 staining, arrows), increased numbers of neutrophils in the epidermis (esterase staining) and macrophages in the dermis (esterase- and F4/80-staining, arrows). **b, c**, H&E staining from affected mouse toes demonstrating granulocytic infiltrates into the joint region (**b**, higher magnification right panel), bone destruction (**c**, arrows) and prominent periostitis (**c**, bottom right panel).

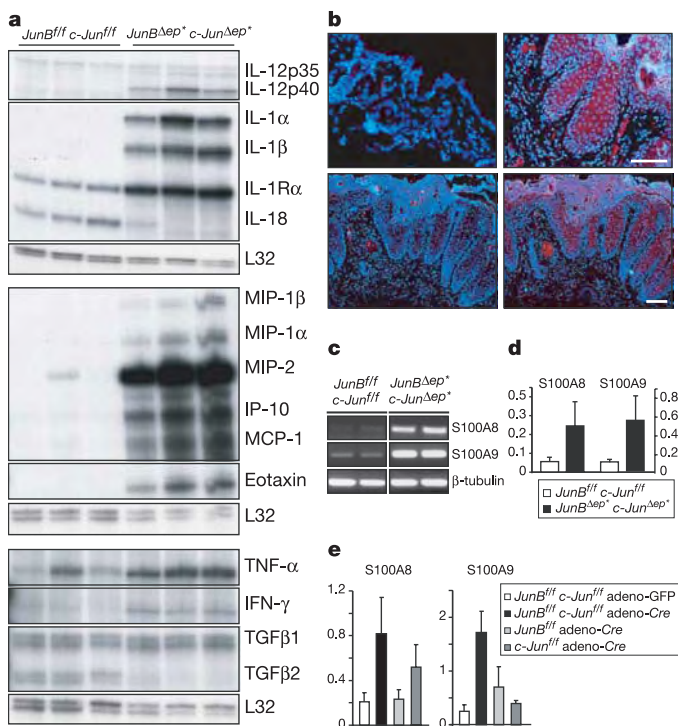


Figure 4 | Mice lacking *JunB* and *c-Jun* in the epidermis exhibit deregulated cytokine expression as observed in psoriasis. **a**, RNase protection assay of epidermal RNA of mice for cytokines, chemokines and their receptors reported to be involved in aspects of psoriasis, before (left) and after (right) inducible deletion of *JunB* and *c-Jun* by tamoxifen injection. Elevated expression of IL-12p40/p35, IL-1α, IL-1β, IL-1Rα, MIP-1α, MIP-1β, MIP-2, IP-10, MCP-1, Eotaxin, TNF-α and IFN-γ detected in diseased mouse mutant epidermis. In contrast, IL-18 and TGF-β2 are downregulated after deletion of *JunB* and *c-Jun*. **b**, Immunohistochemistry of ear sections from wild type (top left panel) and mutant (right panels) mice using anti-S100A8 antibody (serum control, bottom left panel). Upper panels are $\times 40$ and lower panels $\times 20$ magnification. Scale bar, 50 μ m. **c**, **d**, RT-PCR (**c**) and quantification by light cycler RT-PCR (**d**) of S100A8 and S100A9 mRNA. Histologically inconspicuous epidermis isolated 24 h after tamoxifen injection for three consecutive days demonstrated increased expression early in the pathogenesis of psoriasis. **e**, Quantification by light cycler RT-PCR of S100A8 and S100A9 on deletion of *JunB*, *c-Jun* or both after adeno-Cre infections, demonstrating strongly increased expression of S100A8 and S100A9 in double-mutant keratinocytes. Error bars represent the standard deviation.

(data not shown). An RNase protection assay with isolated epidermal RNA demonstrated efficient reduction of mRNA for *JunB* and less efficient for *c-Jun*, whereas other AP-1 mRNAs remained unchanged (see Supplementary Fig. 2a). None of the deregulated genes observed in the diseased mutant epidermis were significantly changed when compared to control (see Supplementary Fig. 2b–d). However, quantification of the epidermal expression of the chemotactic proteins S100A8 and S100A9, by polymerase chain reaction with reverse transcription (RT-PCR) at this early time point showed significantly increased expression in pre-diseased skin (Fig. 4c, d). To further substantiate these findings we used cultured primary keratinocytes and *Cre*-mediated deletion using adenovirus infection. The deletion of the *Jun* proteins resulted in strongly increased expression of S100A8 and S100A9 indicating negative regulation of both S100 genes (Fig. 4e). Efficient induction of S100A8 and S100A9 in cultured keratinocytes required deletion of both *JunB* and *c-Jun* (Fig. 4e), as single knockout keratinocytes showed only slightly increased levels of S100A8 and S100A9 expression. This is in agreement with the *in vivo* data demonstrating increased levels of S100A8 and S100A9 in double-knockout, but not in single-knockout mice, even 18 days after five tamoxifen injections (data not shown). Thus, deregulation

of these proteins appears to be an early molecular event in the development of psoriasis.

Role of T cells in disease development

A functional contribution of T cells in the aetiology of psoriasis is strongly inferred from the presence of T cells in lesioned skin, but also from the induction of psoriasis-like lesions in severe combined immunodeficiency (SCID) mice transplanted with skin of psoriasis patients after transfer of autologous immune cells. The beneficial response to immunosuppressive drugs further supports this notion, however, it is still controversial whether T cells are the cause or consequence of the development of psoriasis²¹. To test the contribution of T cells to the development and progression of psoriasis in our genetic model, we induced deletion of *JunB* and *c-Jun* in mice deficient for Rag2 by applying the protocol shown in Fig. 2b. The skin phenotype of Rag2-deficient *JunB/c-Jun* double-mutant mice was milder when compared with *JunB/c-Jun* double-mutant mice (Fig. 5a). However, epidermal thickening, altered keratinocyte differentiation, vascular dilatation and epidermal micro-abscesses were seen, suggesting a minor role for T and B cells in the aetiology of the disease (Fig. 5b). Consistent with the proposed function in the aetiology of the disease we found strong upregulation of S100A8 and S100A9 in the epidermis of Rag2-deficient mice lacking *JunB* and *c-Jun* (Fig. 5c, d). Notably, these mice showed a similar chemokine/cytokine profile as seen in *JunB^{Δep*} c-Jun^{Δep*}* mice, indicating that T cells are not essential in establishing the chemokine/cytokine profile seen in psoriasis (Fig. 5e). In addition, the paws of mutant mice were almost unaffected and the inflammation of the joint regions was strongly reduced. Bone destruction was absent demonstrating that a functional contribution of immune cells was a prerequisite for the development of arthritic lesions (Fig. 5b).

Development of psoriatic arthritis requires TNF signalling

Recently developed biological agents reported to improve psoriasis and psoriatic arthritis are directed towards inhibiting TNF-α signalling²². To analyse the contribution of TNF-α signalling through TNFR1 in the progression of psoriasis and psoriatic arthritis, we induced deletion of *JunB* and *c-Jun* in mice deficient for TNFR1. Deletion of TNFR1 in *JunB/c-Jun* double-mutant mice could not prevent the development of the skin phenotype, although histological analyses showed a significantly milder phenotype when compared to *JunB/c-Jun* double-mutant mice (Fig. 5a, b). Expression of S100A8 and S100A9 was strongly upregulated (Fig. 5c, d) and a chemokine/cytokine profile comparable to *JunB^{Δep*} c-Jun^{Δep*}* mice was observed, suggesting that TNFR1 signalling is dispensable for the aetiology of the disease (Fig. 5e). However, the inflammation of the joint regions was again almost absent demonstrating a functional contribution of TNF-α signalling through TNFR1 to the aetiology of the arthritic lesions. These results provide strong genetic evidence that the development of the skin phenotype does not fully depend on T-cells or TNF-α signalling through TNFR1, although both are contributors to the disease development and severity.

Discussion

Research into the pathogenesis of psoriasis has been hampered by the lack of an animal disease resembling this common human skin disorder¹. Previous attempts to faithfully reproduce the psoriatic phenotype through expression of inflammatory mediators or keratinocyte growth factors like amphiregulin²³, TNF-α (ref. 24), IL-1α (ref. 25), IFN-γ (ref. 26), KGF²⁷, VEGF²⁸, TGF-β1 (ref. 29), Stat3 (ref. 30) and others²⁸ gave rise to phenotypes with only partial resemblance to psoriasis. Moreover, almost all mouse models discussed above showed no arthritic lesions, although these are present in up to 40% of patients with psoriasis.

Many of the histological and molecular hallmarks of psoriatic skin lesions and arthritis are strikingly reproduced in the animals with deletions of *JunB* and *c-Jun* in epidermal cells. The phenotype that

rapidly develops with 100% penetrance closely resembles the clinical and histological picture of human psoriasis. The disease is prominent in hairless areas of the skin (ears, paws and tail), but is also present in hairy, back skin and the symmetrical distribution is reminiscent of that of psoriatic lesions in humans. Arthritic lesions strongly reminiscent of psoriatic arthritis characterized by inflammatory infiltrates along with massive bone destructions were observed with 100%

penetrance. The diseased epidermis mirrored the changes in the cytokine and chemokine network reported for psoriasis. Expression profiling revealed further evidence that the change in gene expression resembled most of the documented genetic changes described for psoriasis. As in the mouse model an inverse expression pattern for the chemotactic proteins S100A8 and S100A9 (ref. 16) and JunB was found in unaffected and affected areas of skin biopsies from psoriatic

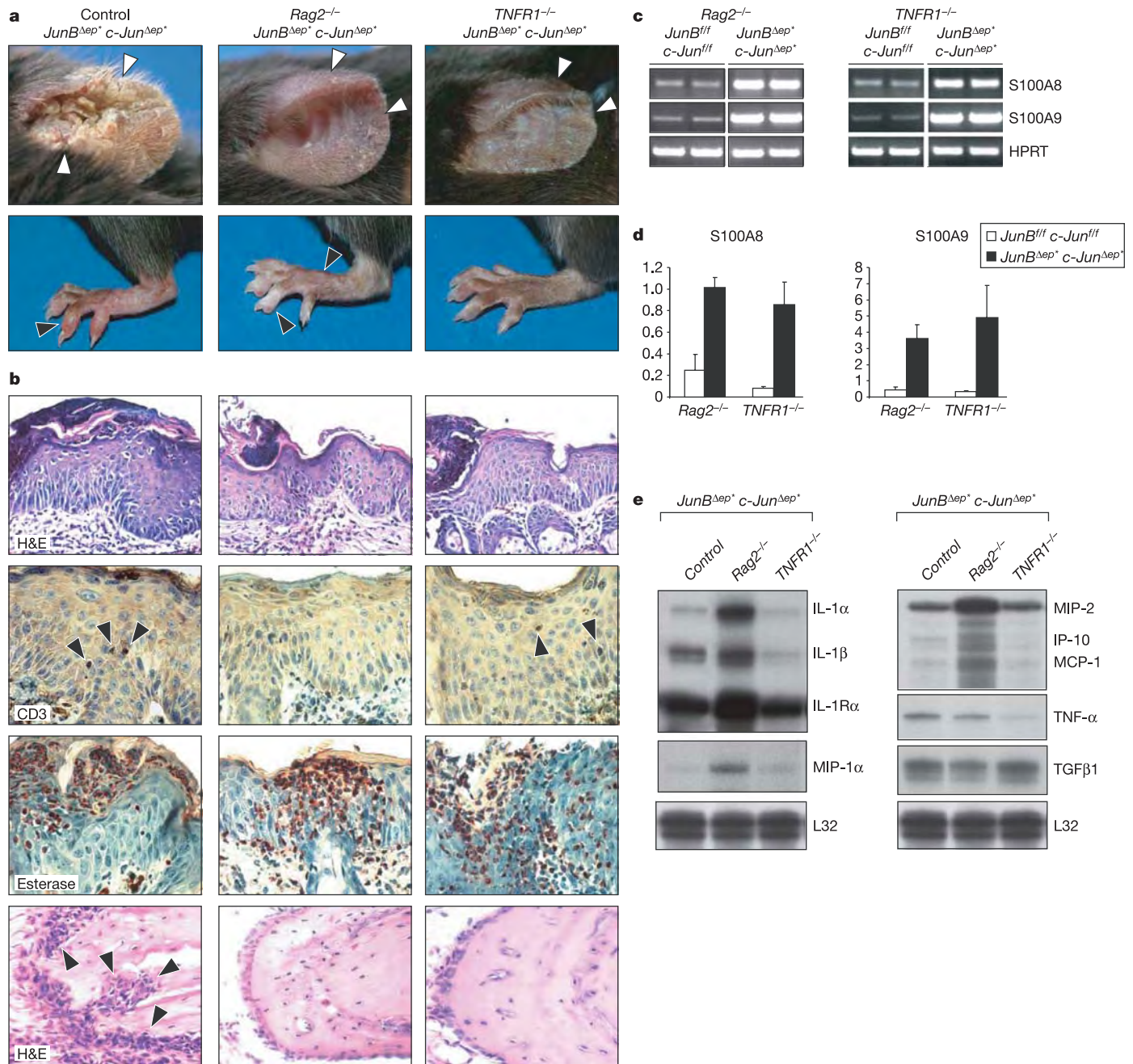


Figure 5 | Inducible deletion of *JunB* and *c-Jun* in *Rag2*- and *TNFR1*-deficient mice. **a**, Macroscopic views of ear and paws from mice 2 weeks after inducible deletion of both *JunB* and *c-Jun* in the epidermis. *Rag2*-deficient mice (middle panels) develop significant skin alterations, but show reduced inflammation of paws (arrow heads). *TNFR1*-deficient mice (right panels) also show skin phenotypes but no signs of paw inflammation. **b**, Histological sections of affected ears and joint regions from mice described in **a**. *Rag2*- and *TNFR1*-deficient mice show reduced hyperproliferation (H&E, $\times 20$) but comparable infiltration of neutrophils (esterase staining, $\times 40$). Note that the specific staining for CD3 ($\times 40$, arrows) is absent in *Rag2*-deficient mice confirming ablation of T cells. No inflammatory infiltrates and bone destruction are seen in the joint regions (bottom panels, $\times 40$) of *Rag2*- or

TNFR1-deficient mice lacking *JunB* and *c-Jun* in the epidermis, indicating the functional contribution of specific immune cells and *TNFR1*-signalling to the aetiology of psoriatic arthritis. Arrows indicate bone destruction in mice lacking *JunB* and *c-Jun* in the epidermis of controls. **c**, **d**, RT-PCR (**c**) and quantification by light cycle RT-PCR of S100A8 and S100A9 (**d**) in diseased *Rag2*- and *TNFR1*-deficient mice lacking *JunB/c-Jun*, demonstrating increased expression of both genes on deletion of *Jun* proteins. **e**, RNase protection assay of epidermal RNA of diseased *Rag2*- and *TNFR1*-deficient mice lacking *JunB* and *c-Jun*, revealing a similar chemokine/cytokine profile as seen in controls. Error bars represent the standard deviation.

patients. These proteins are considered to be potential mediators in psoriasis, because their genes are located in the psoriasis susceptibility region *PSORS4* (1q12; ref. 31). Moreover, the increased expression of the chemotactic proteins S100A8 and S100A9 was identified as an early event in the development of the phenotype, well before any histological alterations or deregulation of other cytokines was observed (Fig. 4c). In addition, strongly increased expression of S100A8 and S100A9 was seen on deletion of *JunB* and *c-Jun* in cultured keratinocytes, suggesting regulation of both genes by Jun proteins (Fig. 4e). Both proteins are potent stimulators of neutrophils³² and passive immunization against S100A8 and S100A9 inhibited neutrophil migration in response to lipopolysaccharide (LPS) injection³³. Therefore, induced chemotaxis of neutrophils by S100A8 and S100A9 on deletion of *JunB* and *c-Jun* can be envisaged as a decisive early event in the establishment of the psoriatic phenotype.

It is interesting to note that in humans, downregulation of JunB alone in the epidermis appears to be the critical factor in the aetiology of psoriasis, whereas in mice, the deletion of both AP-1 components are necessary. This result may be explained by different expression patterns of AP-1 components in the skin of human and mouse, indicating both overlaps and differences in the biological functions of distinct AP-1 proteins³⁴. Reduced AP-1 binding activity was recently reported in lesional skin from psoriatic patients³⁵. However, it is currently unknown whether the reduced epidermal expression of JunB in psoriasis is caused by transcriptional or post-transcriptional regulation. Preliminary experiments using light-cycler RT-PCR with RNA from patients showed no significant changes in RNA levels between affected and unaffected epidermis (data not shown). This may imply that the regulation occurs at the level of altered protein turnover, a mechanism that was recently described for JunB in T cells^{36,37}.

It has been discussed that certain forms of psoriasis, such as the juvenile form, can be triggered by bacterial infections²¹. Therefore, we tested whether treatment of mutant mice with the broad-spectrum antibiotic ciprofloxacin affects the development of the phenotype. Interestingly, preliminary data show that ciprofloxacin significantly delayed the onset of the disease and it seems that the arthritic phenotype did not develop (see Supplementary Fig. 4). This implies that resident bacteria might be responsible for the manifestation of the joint disease in our mouse model.

For several years it has been discussed whether psoriasis represents a fundamental disorder of the skin or the immune system. The presence of T cells does not necessarily implicate a critical participation in the aetiology of the disease and no experiments have been performed in which skin of a healthy individual was forced to exhibit a psoriatic phenotype by transplantation of only T cells^{38,39}. Moreover, it is known that other cell types react in response to so-called T-cell-specific drugs. Cyclosporin A for example has effects on neutrophils⁴⁰ and can inhibit keratinocyte proliferation induced by epidermal growth factor (EGF), TGF- α or IL-6 (ref. 41). Another argument for a functional role of T cells is the association of psoriasis with HLA antigens such as Cw6. Three psoriasis-associated susceptibility alleles have been identified within the *PSORS1* locus (HLA6w, HCR*WWCC and CDSN*5). However, transgenic mice expressing these genes under the control of the keratin 14 (K14) promoter appear phenotypically normal and their skin was histologically indistinguishable from wild-type mice⁴². Moreover, although psoriasis is thought to be a Th1-induced disorder, it develops in human immunodeficiency virus (HIV) infected patients with the same frequency as in the general population. In fact, the phenotype in patients with pre-existing psoriasis may even be worse and more difficult to treat in the presence of HIV disease. This paradoxical exacerbation of psoriasis in AIDS sufferers has not yet been fully explained⁴³.

Strong genetic evidence against an absolute requirement of T cells in the initial development of psoriasis is provided by the deletion of *JunB* and *c-Jun* in a *Rag2*-deficient background. Under these

conditions, histological analyses revealed all the hallmarks of psoriasis, although the skin phenotype was milder when compared to T-cell-competent controls. In addition, a similar chemokine/cytokine profile was observed in *Rag2*-deficient mice lacking JunB and *c-Jun*, indicating that T cells are of minor importance in the development of the disease. We envisage that the role of T cells could be to amplify the initial inflammatory response. Interestingly, only small epidermal lesions at the paws were observed and the inflammation in the joint region was strongly reduced indicating a central role of T cells in the pathogenesis of the arthritic lesions.

Cellular proliferation and deregulated cytokine expression are thought to be central in the pathogenesis of psoriasis. JunB/AP-1 is a well known regulator of cytokine production and cellular proliferation. JunB-deficient mouse fibroblasts are hyperproliferative⁴⁴ and exhibit high basal level expression of cytokines, such as keratinocyte growth factor (KGF) and granulocyte-macrophage colony-stimulating factor (GM-CSF), which promote keratinocyte proliferation in a paracrine fashion⁴⁵. Moreover, downregulation of *JunB* in the myeloid compartment caused a chronic myeloid leukaemia (CML)-like disease^{46,47}. The findings described in this study indicate that epidermis-specific modulation of JunB in humans and of JunB/*c-Jun* in mice induces cytokines/chemokines, which are known to recruit inflammatory cells, thereby contributing to the establishment of the clinical and molecular features observed in psoriasis and psoriatic arthritis. We further provided genetic evidence that the skin disease in mice can develop independently of T cells and that the psoriatic arthritis is primarily mediated by TNF signalling. These data strongly support the recently approved anti-TNFR antibodies as novel therapeutics for psoriasis and psoriatic arthritis.

In summary, we strongly believe that this model with its inducible, high-efficiency and rapid development of the phenotype will be highly suitable for future pre-clinical studies ultimately aimed at understanding and curing this important disease.

METHODS

Generation of *JunB*^{Δep} *c-Jun*^{Δep} mice. Mice carrying a floxed *JunB* allele⁴⁸ (*JunB*^{flf}) and/or a floxed *c-Jun* allele⁴⁹ (*c-Jun*^{flf}) were crossed to transgenic mice expressing the Cre recombinase-estrogen receptor fusion under the control of the keratinocyte-specific keratin 5 promoter⁶ (K5-Cre-ER^T) to obtain *JunB*^{flf}, *c-Jun*^{flf} and *JunB*^{flf} *c-Jun*^{flf} K5-Cre-ER^T mice.

Inducible deletion of *JunB* and *c-Jun*. Eight-week-old mutant mice and their littermate controls were injected daily (intraperitoneal), five times with 1 mg tamoxifen⁵⁰ (Sigma) or three times with 1 mg tamoxifen (shortened protocol) to obtain epidermis-specific, single- and double-knockout mice.

Adeno-Cre mediated deletion *in vitro*. Primary keratinocytes were infected (300 particles per cell) with adeno-Cre or adeno-green fluorescent protein (GFP; a gift from M. Cotton) in medium with reduced serum (4% chelated fetal calf serum). After 2 h, normal keratinocyte medium (8% chelated FCS) was added, and after 48 h cells were collected for further analyses.

RNase protection assay and RT-PCR. Epidermis was separated from dermis by trypsin digestion (0.25%) for 1 h at 37 °C. RNase protection assays were performed using the RiboQuant multi-probe RNase protection assay systems mJun/Fos, mck2, mck3 and mck5c (PharMingen) according to the manufacturer's instructions. Standard real-time PCR was performed with the following primers: tubulin-up 5'-CAACGTCAGACGGCCGTGTG-3'; tubulin-down 5'-ACTGGCGGGGTGTAGGTAAAGGTG-3'; S100A8up 5'-GGAATCACC ATGCCCTCTACA-3'; S100A8down 5'-TGCCACGCCACCCCTTATC-3'; S100A9up 5'-GAGCGCAGCATAACCACCATC-3'; S100A9down 5'-AGCCATT CCCTTACACTTG-3'; hJunB-up 5'-GCAGGTGGCCAGCTCAAACAG-3'; hJunB-down 5'-GCCGCGATCGCCCCCTCTT-3'; hc-Jun-up 5'-TTGCGGCCCC CGAACTTGTC-3'; hc-Jun-down 5'-CTCGGCGAACCCCTCTGCTCAT-3'.

Immunohistochemistry. Human skin biopsies (*n* = 8, obtained for diagnostic purposes from patients suffering from psoriasis) and mouse tissues were fixed overnight with neutral buffered 4% paraformaldehyde (PFA) at 4 °C and either directly paraffin embedded or after decalcification (bone) in 0.5% EDTA for 12 d. Five-micrometre sections were stained either with hematoxylin and eosin (H&E) or processed further. Staining for anti JunB, CD31, CD3 and F4/80 (Santa Cruz) was performed after antigen-retrieval (Dako; S1699) with the MultiLink Dako system (Dako; E0453) according to the manufacturer's instructions. Expression of S100A8 was detected by immunofluorescence using polyclonal

goat antibodies (Santa Cruz; SC 8112) and secondary donkey-anti-goat-Cy3 antibody (Dianova). For nuclear staining H33342 (Calbiochem) was added in a final concentration of $1 \mu\text{g ml}^{-1}$ to the secondary antibody dilution. To quantify immunofluorescence signals of JunB and c-Jun in different layers of normal and psoriatic skin the intensity of 30 nuclei of every layer was measured using Metamorph Software. After normalization for the intensity of background staining in every epidermal layer statistical analysis was performed using two-sided *t*-test, where $P < 0.05$ was accepted as significant; error bars represent the standard deviation.

cDNA libraries and processing of microarrays. Global gene expression profiling was performed on two different microarrays, comprising: (1) the murine 20 k ArrayTAG (ref. 12; LION Bioscience); and (2) the 22.4 k NIA cDNA collection from the National Institute of Ageing, combining the 15 k NIA clone set^{13,14} extended by 7.4 k additional sequences¹⁵. Labelling of the probes, hybridization of samples and analysis of fluorescent microarray images was performed as described previously¹⁴. To account for the systemic error caused by different properties of the fluorescent dyes concerning incorporation, mean brightness and background noise, the hybridizations were performed in duplicate with reversed assignment of fluorescent dyes ('colour switch').

Received 31 March; accepted 25 May 2005.

- Schon, M. P. Animal models of psoriasis—what can we learn from them? *J. Invest. Dermatol.* **112**, 405–410 (1999).
- Nickoloff, B. J. & Nestle, F. O. Recent insights into the immunopathogenesis of psoriasis provide new therapeutic opportunities. *J. Clin. Invest.* **113**, 1664–1675 (2004).
- Shaulian, E. & Karin, M. AP-1 as a regulator of cell life and death. *Nature Cell Biol.* **4**, E131–E136 (2002).
- Eferl, R. & Wagner, E. F. AP-1: a double-edged sword in tumorigenesis. *Nature Rev. Cancer* **3**, 859–868 (2003).
- Mehic, D., Bakiri, L., Ghannadan, M., Wagner, E. F. & Tschachler, E. Fos and jun proteins are specifically expressed during differentiation of human keratinocytes. *J. Invest. Dermatol.* **124**, 212–220 (2005).
- Brocard, J. et al. Spatio-temporally controlled site-specific somatic mutagenesis in the mouse. *Proc. Natl Acad. Sci. USA* **94**, 14559–14563 (1997).
- Bonifati, C. & Ameglio, F. Cytokines in psoriasis. *Int. J. Dermatol.* **38**, 241–251 (1999).
- Uyemura, K., Yamamura, M., Fivenson, D. F., Modlin, R. L. & Nickoloff, B. J. The cytokine network in lesional and lesion-free psoriatic skin is characterized by a T-helper type 1 cell-mediated response. *J. Invest. Dermatol.* **101**, 701–705 (1993).
- Kong, J. & Li, Y. C. Upregulation of interleukin-18 expression in mouse primary keratinocytes induced to differentiate by calcium. *Arch. Dermatol. Res.* **294**, 370–376 (2002).
- Lee, E. et al. Increased expression of interleukin 23 p19 and p40 in lesional skin of patients with psoriasis vulgaris. *J. Exp. Med.* **199**, 125–130 (2004).
- Hashimoto, K. Regulation of keratinocyte function by growth factors. *J. Dermatol. Sci.* **24**, S46–S50 (2000).
- Schlingemann, J. et al. Profile of gene expression induced by the tumour promoter TPA in murine epithelial cells. *Int. J. Cancer* **104**, 699–708 (2003).
- Tanaka, T. S. et al. Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. *Proc. Natl Acad. Sci. USA* **97**, 9127–9132 (2000).
- Florin, L. et al. Identification of novel AP-1 target genes in fibroblasts regulated during cutaneous wound healing. *Oncogene* **23**, 7005–7017 (2004).
- VanBuren, V. et al. Assembly, verification and initial annotation of the NIA mouse 7.4K cDNA clone set. *Genome Res.* **12**, 1999–2003 (2002).
- Broome, A. M., Ryan, D. & Eckert, R. L. S100 protein subcellular localization during epidermal differentiation and psoriasis. *J. Histochem. Cytochem.* **51**, 675–685 (2003).
- Waseem, A. et al. Keratin 15 expression in stratified epithelia: downregulation in activated keratinocytes. *J. Invest. Dermatol.* **112**, 362–369 (1999).
- Campbell, L. et al. Downregulation and altered spatial pattern of caveolin-1 in chronic plaque psoriasis. *Br. J. Dermatol.* **147**, 701–709 (2002).
- Apte, S. S., Olsen, B. R. & Murphy, G. The gene structure of tissue inhibitor of metalloproteinases (TIMP)-3 and its inhibitory activities define the distinct TIMP gene family. *J. Biol. Chem.* **270**, 14313–14318 (1995).
- Black, R. A. TIMP3 checks inflammation. *Nature Genet.* **36**, 934–935 (2004).
- Nickoloff, B. J. et al. Is psoriasis a T-cell disease? *Exp. Dermatol.* **9**, 359–375 (2000).
- Galadari, H., Fuchs, B. & Lebowitz, M. Newly available treatments for psoriatic arthritis and their impact on skin psoriasis. *Int. J. Dermatol.* **42**, 231–237 (2003).
- Cook, P. W., Pittelkow, M. R. & Piepkorn, M. Overexpression of amphiregulin in the epidermis of transgenic mice induces a psoriasis-like cutaneous phenotype. *J. Invest. Dermatol.* **113**, 860 (1999).
- Cheng, J. et al. Cachexia and graft-vs.-host-disease-type skin changes in keratin promoter-driven TNF alpha transgenic mice. *Genes Dev.* **6**, 1444–1456 (1992).
- Groves, R. W., Mizutani, H., Kieffer, J. D. & Kupper, T. S. Inflammatory skin disease in transgenic mice that express high levels of interleukin 1 α in basal epidermis. *Proc. Natl Acad. Sci. USA* **92**, 11874–11878 (1995).
- Carroll, J. M., Crompton, T., Seery, J. P. & Watt, F. M. Transgenic mice expressing IFN-gamma in the epidermis have eczema, hair hypopigmentation, and hair loss. *J. Invest. Dermatol.* **108**, 412–422 (1997).
- Guo, L., Yu, Q. C. & Fuchs, E. Targeting expression of keratinocyte growth factor to keratinocytes elicits striking changes in epithelial differentiation in transgenic mice. *EMBO J.* **12**, 973–986 (1993).
- Xia, Y. P. et al. Transgenic delivery of VEGF to mouse skin leads to an inflammatory condition resembling human psoriasis. *Blood* **102**, 161–168 (2003).
- Li, A. G., Wang, D., Feng, X. H. & Wang, X. J. Latent TGF β 1 overexpression in keratinocytes results in a severe psoriasis-like skin disorder. *EMBO J.* **23**, 1770–1781 (2004).
- Sano, S. et al. Stat3 links activated keratinocytes and immunocytes required for development of psoriasis in a novel transgenic mouse model. *Nature Med.* **11**, 43–49 (2005).
- Semprini, S. et al. Evidence for differential S100 gene over-expression in psoriatic patients from genetically heterogeneous pedigrees. *Hum. Genet.* **111**, 310–313 (2002).
- Ryckman, C., Vandal, K., Rouleau, P., Talbot, M. & Tessier, P. A. Proinflammatory activities of S100: proteins S100A8, S100A9, and S100A8/A9 induce neutrophil chemotaxis and adhesion. *J. Immunol.* **170**, 3233–3242 (2003).
- Vandal, K. et al. Blockade of S100A8 and S100A9 suppresses neutrophil migration in response to lipopolysaccharide. *J. Immunol.* **171**, 2602–2609 (2003).
- Angel, P., Szabowski, A. & Schorpp-Kistner, M. Function and regulation of AP-1 subunits in skin physiology and pathology. *Oncogene* **20**, 2413–2423 (2001).
- Johansen, C., Kragballe, K., Rasmussen, M., Dam, T. N. & Iversen, L. Activator protein 1 DNA binding activity is decreased in lesional psoriatic skin compared with nonlesional psoriatic skin. *Br. J. Dermatol.* **151**, 600–607 (2004).
- Fang, D. et al. Dysregulation of T lymphocyte function in itchy mice: a role for Itch in TH2 differentiation. *Nature Immunol.* **3**, 281–287 (2002).
- Gao, M. et al. Jun turnover is controlled through JNK-dependent phosphorylation of the E3 ligase Itch. *Science* **306**, 271–275 (2004).
- Nickoloff, B. J. The search for pathogenic T cells and the genetic basis of psoriasis using a severe combined immunodeficient mouse model. *Cutis* **65**, 110–114 (2000).
- Boyman, O. et al. Spontaneous development of psoriasis in a new animal model shows an essential role for resident T cells and tumour necrosis factor- α . *J. Exp. Med.* **199**, 731–736 (2004).
- Janco, R. L. & English, D. Cyclosporine and human neutrophil function. *Transplantation* **35**, 501–503 (1983).
- Karashima, T., Hachisuka, H. & Sasai, Y. FK506 and cyclosporin A inhibit growth factor-stimulated human keratinocyte proliferation by blocking cells in the G0/G1 phases of the cell cycle. *J. Dermatol. Sci.* **12**, 246–254 (1996).
- Elomaa, O. et al. Transgenic mouse models support HCR as an effector gene in the PSORS1 locus. *Hum. Mol. Genet.* **13**, 1551–1561 (2004).
- Namazi, M. R. Paradoxical exacerbation of psoriasis in AIDS: proposed explanations including the potential roles of substance P and gram-negative bacteria. *Autoimmunity* **37**, 67–71 (2004).
- Passague, E. & Wagner, E. F. JunB suppresses cell proliferation by transcriptional activation of p16^{INK4a} expression. *EMBO J.* **19**, 2969–2979 (2000).
- Szabowski, A. et al. c-Jun and JunB antagonistically control cytokine-regulated mesenchymal-epidermal interaction in skin. *Cell* **103**, 745–755 (2000).
- Passague, E., Jochum, W., Schorpp-Kistner, M., Mohle-Steinlein, U. & Wagner, E. F. Chronic myeloid leukemia with increased granulocyte progenitors in mice lacking junB expression in the myeloid lineage. *Cell* **104**, 21–32 (2001).
- Passague, E., Wagner, E. F. & Weissman, I. L. JunB deficiency leads to a myeloproliferative disorder arising from hematopoietic stem cells. *Cell* **119**, 431–443 (2004).
- Kenner, L. et al. Mice lacking JunB are osteopenic due to cell-autonomous osteoblast and osteoclast defects. *J. Cell Biol.* **164**, 613–623 (2004).
- Behrens, A. et al. Impaired postnatal hepatocyte proliferation and liver regeneration in mice lacking c-jun in the liver. *EMBO J.* **21**, 1782–1790 (2002).
- Vasioukhin, V., Degenstein, L., Wise, B. & Fuchs, E. The magical touch: genome targeting in epidermal stem cells induced by tamoxifen application to mouse skin. *Proc. Natl Acad. Sci. USA* **96**, 8551–8556 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to M. Sibilia, G. Stingl, D. Maurer, J. Smolen, G. Schett and A. Rot for critical comments and suggestions to the manuscript, M. Cotton for providing adeno-Cre viruses, H. Tkadletz for help in preparing the illustrations and J. Hess for support with S100 protein expression analyses. The IMP is funded by Boehringer Ingelheim and this work was supported by grants from the Austrian Research Foundation, the Deutsche Forschungsgemeinschaft and by the Research Training Network (RTN) Program of the European Community.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.F.W. (wagner@imp.univie.ac.at).

ARTICLES

Genome sequencing in microfabricated high-density picolitre reactors

Marcel Margulies^{1*}, Michael Egholm^{1*}, William E. Altman¹, Said Attiya¹, Joel S. Bader¹, Lisa A. Bemben¹, Jan Berka¹, Michael S. Braverman¹, Yi-Ju Chen¹, Zhoutao Chen¹, Scott B. Dewell¹, Lei Du¹, Joseph M. Fierro¹, Xavier V. Gomes¹, Brian C. Godwin¹, Wen He¹, Scott Helgesen¹, Chun He Ho¹, Gerard P. Irzyk¹, Szilveszter C. Jando¹, Maria L. I. Alenquer¹, Thomas P. Jarvie¹, Kshama B. Jirage¹, Jong-Bum Kim¹, James R. Knight¹, Janna R. Lanza¹, John H. Leamon¹, Steven M. Lefkowitz¹, Ming Lei¹, Jing Li¹, Kenton L. Lohman¹, Hong Lu¹, Vinod B. Makhijani¹, Keith E. McDade¹, Michael P. McKenna¹, Eugene W. Myers², Elizabeth Nickerson¹, John R. Nobile¹, Ramona Plant¹, Bernard P. Puc¹, Michael T. Ronan¹, George T. Roth¹, Gary J. Sarkis¹, Jan Fredrik Simons¹, John W. Simpson¹, Maithreyan Srinivasan¹, Karrie R. Tartaro¹, Alexander Tomasz³, Kari A. Vogt¹, Greg A. Volkmer¹, Shally H. Wang¹, Yong Wang¹, Michael P. Weiner⁴, Pengguang Yu¹, Richard F. Begley¹ & Jonathan M. Rothberg¹

The proliferation of large-scale DNA-sequencing projects in recent years has driven a search for alternative methods to reduce time and cost. Here we describe a scalable, highly parallel sequencing system with raw throughput significantly greater than that of state-of-the-art capillary electrophoresis instruments. The apparatus uses a novel fibre-optic slide of individual wells and is able to sequence 25 million bases, at 99% or better accuracy, in one four-hour run. To achieve an approximately 100-fold increase in throughput over current Sanger sequencing technology, we have developed an emulsion method for DNA amplification and an instrument for sequencing by synthesis using a pyrosequencing protocol optimized for solid support and picolitre-scale volumes. Here we show the utility, throughput, accuracy and robustness of this system by shotgun sequencing and *de novo* assembly of the *Mycoplasma genitalium* genome with 96% coverage at 99.96% accuracy in one run of the machine.

DNA sequencing has markedly changed the nature of biomedical research and medicine. Reductions in the cost, complexity and time required to sequence large amounts of DNA, including improvements in the ability to sequence bacterial and eukaryotic genomes, will have significant scientific, economic and cultural impact. Large-scale sequencing projects, including whole-genome sequencing, have usually required the cloning of DNA fragments into bacterial vectors, amplification and purification of individual templates, followed by Sanger sequencing¹ using fluorescent chain-terminating nucleotide analogues² and either slab gel or capillary electrophoresis. Current estimates put the cost of sequencing a human genome between \$10 million and \$25 million³. Alternative sequencing methods have been described^{4–8}; however, no technology has displaced the use of bacterial vectors and Sanger sequencing as the main generators of sequence information.

Here we describe an integrated system whose throughput routinely enables applications requiring millions of bases of sequence information, including whole-genome sequencing. Our focus has been on the co-development of an emulsion-based method^{9–11} to isolate and amplify DNA fragments *in vitro*, and of a fabricated substrate and instrument that performs pyrophosphate-based sequencing (pyrosequencing^{5,12}) in picolitre-sized wells.

In a typical run we generate over 25 million bases with a Phred quality score of 20 or better (predicted to have an accuracy of 99% or higher). Although this Phred 20 quality throughput is significantly

higher than that of Sanger sequencing by capillary electrophoresis, it is currently at the cost of substantially shorter reads and lower average individual read accuracy. Sanger-based capillary electrophoresis sequencing systems produce up to 700 bases of sequence information from each of 96 DNA templates at an average read accuracy of 99.4% in 1 h, or 67,000 bases per hour, with substantially all of the bases having Phred 20 or better quality²³. We further characterize the performance of the system and demonstrate that it is possible to assemble bacterial genomes *de novo* from relatively short reads by sequencing a known bacterial genome, *Mycoplasma genitalium* (580,069 bases), and comparing our shotgun sequencing and *de novo* assembly with the results originally obtained for this genome¹³. The results of shotgun sequencing and *de novo* assembly of a larger bacterial genome, that of *Streptococcus pneumoniae*¹⁴ (2.1 megabases (Mb)), are presented in Supplementary Table 4.

Emulsion-based sample preparation

We generate random libraries of DNA fragments by shearing an entire genome and isolating single DNA molecules by limiting dilution (Supplementary Methods). Specifically, we randomly fragment the entire genome, add specialized common adapters to the fragments, capture the individual fragments on their own beads and, within the droplets of an emulsion, clonally amplify the individual fragment (Fig. 1a, b). Unlike in current sequencing technology, our approach does not require subcloning in bacteria or the handling of

¹454 Life Sciences Corp., 20 Commercial Street, Branford, Connecticut 06405, USA. ²University of California, Berkeley, California 94720, USA. ³Laboratory of Microbiology, The Rockefeller University, New York, New York 10021, USA. ⁴The Rothberg Institute for Childhood Diseases, 530 Whitfield Street, Guilford, Connecticut 06437, USA.

*These authors contributed equally to this work.

individual clones; the templates are handled in bulk within the emulsions^{9–11}.

Sequencing in fabricated picolitre-sized reaction vessels

We perform sequencing by synthesis simultaneously in open wells of a fibre-optic slide using a modified pyrosequencing protocol that is designed to take advantage of the small scale of the wells. The fibre-optic slides are manufactured by slicing of a fibre-optic block that is obtained by repeated drawing and fusing of optic fibres. At each iteration, the diameters of the individual fibres decrease as they are hexagonally packed into bundles of increasing cross-sectional sizes. Each fibre-optic core is 44 μm in diameter and surrounded by 2–3 μm of cladding; etching of each core creates reaction wells approximately 55 μm in depth with a centre-to-centre distance of 50 μm (Fig. 1c), resulting in a calculated well size of 75 pl and a well density of 480 wells mm^{-2} . The slide, containing approximately 1.6 million wells¹⁵, is loaded with beads and mounted in a flow chamber designed to create a 300- μm high channel, above the well openings, through which the sequencing reagents flow (Fig. 2a, b). The unetched base of the slide is in optical contact with a second fibre-optic imaging bundle bonded to a charge-coupled device (CCD)

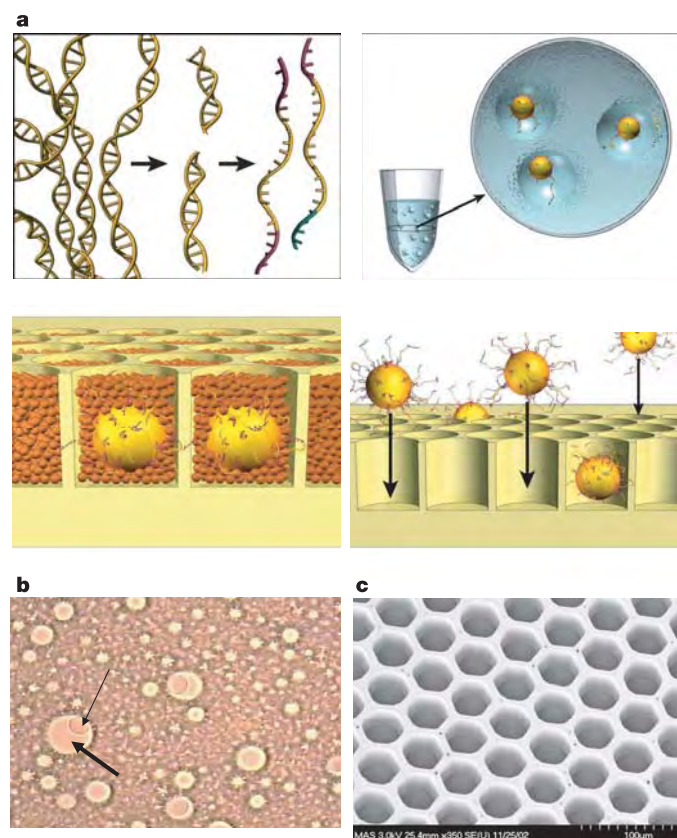


Figure 1 | Sample preparation. **a**, Genomic DNA is isolated, fragmented, ligated to adapters and separated into single strands (top left). Fragments are bound to beads under conditions that favour one fragment per bead, the beads are captured in the droplets of a PCR-reaction-mixture-in-oil emulsion and PCR amplification occurs within each droplet, resulting in beads each carrying ten million copies of a unique DNA template (top right). The emulsion is broken, the DNA strands are denatured, and beads carrying single-stranded DNA clones are deposited into wells of a fibre-optic slide (bottom right). Smaller beads carrying immobilized enzymes required for pyrophosphate sequencing are deposited into each well (bottom left). **b**, Microscope photograph of emulsion showing droplets containing a bead and empty droplets. The thin arrow points to a 28- μm bead; the thick arrow points to an approximately 100- μm droplet. **c**, Scanning electron micrograph of a portion of a fibre-optic slide, showing fibre-optic cladding and wells before bead deposition.

sensor, allowing the capture of emitted photons from the bottom of each individual well (Fig. 2c; see also Supplementary Methods).

We developed a three-bead system, and optimized the components to achieve high efficiency on solid support. The combination of picolitre-sized wells, enzyme loading uniformity allowed by the small beads and enhanced solid support chemistry enabled us to develop a method that extends the useful read length of sequencing-by-synthesis to 100 bases (Supplementary Methods).

In the flow chamber cyclically delivered reagents flow perpendicularly to the wells. This configuration allows simultaneous extension reactions on template-carrying beads within the open wells and relies on convective and diffusive transport to control the addition or removal of reagents and by-products. The timescale for diffusion into and out of the wells is on the order of 10 s in the current configuration and is dependent on well depth and flow channel height. The timescales for the signal-generating enzymatic reactions are on the order of 0.02–1.5 s (Supplementary Methods). The current reaction is dominated by mass transport effects, and improvements based on faster delivery of reagents are possible. Well depth was selected on the basis of a number of competing requirements: (1) wells need to be deep enough for the DNA-carrying beads to remain in the wells in the presence of convective transport past the wells; (2) they must be sufficiently deep to provide adequate isolation against diffusion of by-products from a well in which incorporation is taking place to a well where no incorporation is occurring; and (3) they must be shallow enough to allow rapid diffusion of nucleotides into the wells and rapid washing out of remaining nucleotides at the end of each flow cycle to enable high sequencing throughput and reduced reagent use. After the flow of each nucleotide, a wash containing apyrase is used to ensure that nucleotides do not remain in any well before the next nucleotide being introduced.

Base calling of individual reads

Nucleotide incorporation is detected by the associated release of inorganic pyrophosphate and the generation of photons^{5,12}. Wells containing template-carrying beads are identified by detecting a known four-nucleotide 'key' sequence at the beginning of the read

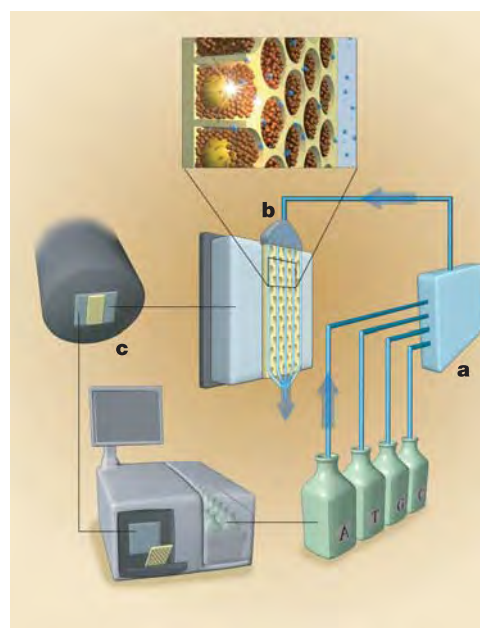


Figure 2 | Sequencing instrument. The sequencing instrument consists of the following major subsystems: a fluidic assembly (**a**), a flow chamber that includes the well-containing fibre-optic slide (**b**), a CCD camera-based imaging assembly (**c**), and a computer that provides the necessary user interface and instrument control.

(Supplementary Methods). Raw signals are background-subtracted, normalized and corrected. The normalized signal intensity at each nucleotide flow, for a particular well, indicates the number of nucleotides, if any, that were incorporated. This linearity in signal is preserved to at least homopolymers of length eight (Supplementary Fig. 6). In sequencing by synthesis a very small number of templates on each bead lose synchronism (that is, either get ahead of, or fall behind, all other templates in sequence¹⁶). The effect is primarily due to leftover nucleotides in a well (creating ‘carry forward’) or to incomplete extension. Typically, we observe a carry forward rate of 1–2% and an incomplete extension rate of 0.1–0.3%. Correction of these shifts is essential because the loss of synchronism is a cumulative effect that degrades the quality of sequencing at longer read lengths. We have developed algorithms, based on detailed models of the underlying physical phenomena, that allow us to determine, and correct for, the amounts of carry forward and incomplete extension occurring in individual wells (Supplementary Methods). Figure 3 shows the processed result, a 113-bases-long read generated in the *M. genitalium* run discussed below. To assess sequencing performance and the effectiveness of the correction algorithms, independently of artefacts introduced during the emulsion-based sample preparation, we created test fragments with difficult-to-sequence stretches of identical bases of increasing length (homopolymers) (Supplementary Methods and Supplementary Fig. 4). Using these test fragments, we have verified that at the individual read level we achieve base call accuracy of approximately 99.4%, at read lengths in excess of 100 bases (Table 1).

High-quality reads and consensus accuracy

Before base calling or aligning reads, we select high-quality reads without relying on a priori knowledge of the genome or template being sequenced (Supplementary Methods). This selection is based on the observation that poor-quality reads have a high proportion of signals that do not allow a clear distinction between a flow during which no nucleotide was incorporated and a flow during which one or more nucleotide was incorporated. When base calling individual reads, errors can occur because of signals that have ambiguous values (Supplementary Fig. 5). To improve the usability of our reads, we also developed a metric that allows us to estimate *ab initio* the quality (or probability of correct base call) of each base of a read, analogous to the Phred score¹⁷ used by current Sanger sequencers (Supplementary Methods and Supplementary Fig. 8).

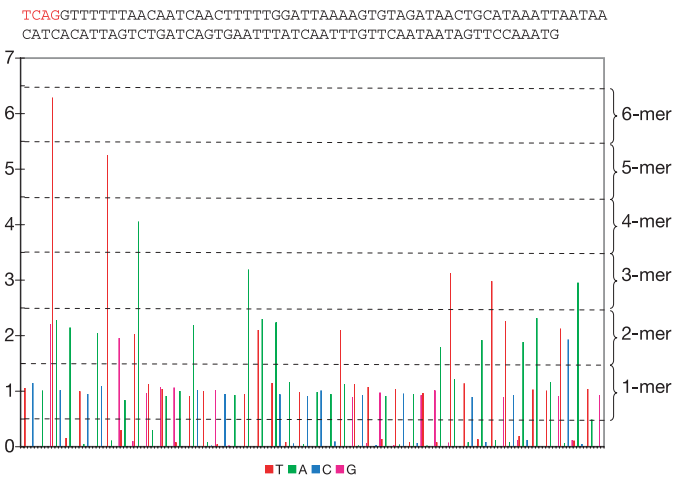


Figure 3 | Flowgram of a 113-bases read from an *M. genitalium* run. Nucleotides are flowed in the order T, A, C, G. The sequence is shown above the flowgram. The signal value intervals corresponding to the various homopolymers are indicated on the right. The first four bases (in red, above the flowgram) constitute the ‘key’ sequence, used to identify wells containing a DNA-carrying bead.

Higher quality sequence can be achieved by taking advantage of the high over sampling that our system affords and building a consensus sequence. Sequences are aligned to one another using the signal strengths at each nucleotide flow, rather than individual base calls, to determine optimal alignment (Supplementary Methods). The corresponding signals are then averaged, after which base calling is performed. This approach greatly improves the accuracy of the sequence (Supplementary Fig. 7) and provides an estimate of the quality of the consensus base. We refer to that quality measure as the Z-score—it is a measure of the spread of signals in all the reads at one location and the distance between the average signal and the closest base-calling threshold value. In both re-sequencing and *de novo* sequencing, as the minimum Z-score is raised the consensus accuracy increases, while coverage decreases; approximately half of the excluded bases, as the Z-score is increased, belong to homopolymers of length four and larger. Sanger sequencers usually require a depth of coverage at any base of three or more in order to achieve a consensus accuracy of 99.99%. To achieve a minimum of threefold coverage of 95% of the unique portions of a typical genome requires approximately seven- to eightfold over sampling. Owing to our higher error rate, we have observed that comparable consensus accuracies, over a similar fraction of a genome, are achieved with a depth of coverage of four or more, requiring approximately ten to twelve times over sampling.

Mycoplasma genitalium

Mycoplasma genomic DNA was fragmented and prepared into a sequencing library as described above. (This was accomplished by a single individual in 4 h.) After emulsion polymerase chain reaction (PCR) and bead deposition onto a 60 × 60 mm² fibre-optic slide, a process which took one individual 6 h, 42 cycles of four nucleotides were flowed through the sequencing system in an automated 4-h run of the instrument. The results are summarized in Table 2. In order to measure the quality of individual reads, we aligned each high quality read to the reference genome at 70% stringency using flow-space mapping and criteria similar to those used previously in assessing the accuracy of other base callers¹⁷. When assessing sequencing quality, only reads that mapped to unique locations in the reference genome were included. Because this process excludes repeat regions (parts of the genome for which corresponding flowgrams are 70% similar to one another), the selected reads did not cover the genome completely. Figure 4a illustrates the distribution of read lengths for this run. The average read length was 110 bases, the resulting over sample 40-fold, and 84,011 reads (27.4%) were perfect. Figure 4b summarizes the average error as a function of base position. Coverage of non-repeat regions was consistent with the sample preparation and emulsion not being biased (Supplementary Fig. 8). At the individual read level, we observe an insertion and deletion error rate of approximately 3.3%; substitution errors have a much lower rate, on the order of 0.5%. When using these reads without any Z-score restriction, we covered 99.94% of the genome in ten contiguous regions with a consensus accuracy of 99.97%. The error rate in homopolymers is significantly reduced in the consensus sequence (Supplementary Fig. 7). Of the bases not covered by this consensus

Table 1 | Summary of sequencing statistics for test fragments

Size of fibre-optic slide	60 × 60 mm ²
Run time/number of cycles	243 min/42
Test fragment reads	497,893
Average read length (bases)	108
Number of bases in test fragments	53,705,267
Bases with a Phred score of 20 and above	47,181,792
Individual read insertion error rate	0.44%
Individual read deletion error rate	0.15%
Individual read substitution error rate	0.004%
All errors	0.60%

sequence (366 bases), all belonged to excluded repeat regions. Setting a minimum Z-score equal to 4, coverage was reduced to 98.1% of the genome, while consensus accuracy increased to 99.996%. We further demonstrated the reproducibility of the system by repeating the whole-genome sequencing of *M. genitalium* an additional eight times, achieving a 40-fold coverage of the genome in each of the eight separate instrument runs (Supplementary Table 3).

We assembled the *M. genitalium* reads from a single run into 25 contigs with an average length of 22.4 kb. One of these contigs was misassembled due to a collapsed tandem repeat region of 60 bases, and was corrected by hand. The original sequencing of *M. genitalium* resulted in 28 contigs before directed sequencing used for finishing the sequence¹³. Our assembly covered 96.54% of the genome and attained a consensus accuracy of 99.96%. Non-resolvable repeat regions amount to 3% of the genome: we therefore covered 99.5% of the unique portions of the genome. Sixteen of the breaks between contigs were due to non-resolvable repeat regions, two were due to missed overlapping reads (our read filter and trimmer are not perfect and the algorithms we use to perform the pattern matching of flowgrams occasionally miss valid overlaps) and the remainder to thin read coverage. Setting a minimum Z-score of 4, coverage was reduced to 95.27% of the genome (98.2% of the resolvable part of the genome) with the consensus accuracy increasing to 99.994%.

Discussion

We have demonstrated the simultaneous acquisition of hundreds of thousands of sequence reads, 80–120 bases long, at 96% average accuracy in a single run of the instrument using a newly developed *in vitro* sample preparation methodology and sequencing technology. With Phred 20 as a cutoff, we show that our instrument is able to produce over 47 million bases from test fragments and 25 million bases from genomic libraries. We used test fragments to de-couple our sample preparation methodology from our sequencing technology. The decrease in single-read accuracy from 99.4% for test fragments to 96% for genomic libraries is primarily due to a lack of clonality in a fraction of the genomic templates in the emulsion, and is not an inherent limitation of the sequencing technology. Most of the remaining errors result from a broadening of signal distributions, particularly for large homopolymers (seven or more), leading to ambiguous base calls. Recent work on the sequencing

chemistry and algorithms that correct for crosstalk between wells suggests that the signal distributions will narrow, with an attendant reduction in errors and increase in read lengths. In preliminary experiments with genomic libraries that also include improvements in the emulsion protocol, we are able to achieve, using 84 cycles, read lengths of 200 bases with accuracies similar to those demonstrated here for 100 bases. On occasion, at 168 cycles, we have generated individual reads that are 100% accurate over greater than 400 bases.

Using *M. genitalium* we demonstrate that short fragments a priori do not prohibit the *de novo* assembly of bacterial genomes. In fact, the larger over sampling afforded by the throughput of our system resulted in a draft sequence having fewer contigs than with Sanger reads, with substantially less effort. By taking advantage of the over sampling, consensus accuracies greater than 99.96% were achieved for this genome. Further quality filtering of the assembly results in the selection of a consensus sequence with accuracy exceeding 99.99% while incurring only a minor loss of genome coverage. Comparable results were seen when we shotgun sequenced and *de novo* assembled the 2.1-Mb genome of *Streptococcus pneumoniae*¹⁴ (Supplementary Table 4). The *de novo* assembly of genomes more complex than bacteria, including mammalian genomes, may require the development of methods similar to those developed for Sanger sequencing, to prepare and sequence paired end libraries that can span repeats in these genome. To facilitate the use of paired end libraries we have developed methods to sequence, in an individual well, from both ends of genomic template, and plan to add paired end read capabilities to our assembler (Supplementary Methods).

Future increases in throughput, and a concomitant reduction in cost per base, may come from the continued miniaturization of the fibre-optic reactors, allowing more sequence to be produced per unit area—a scaling characteristic similar to that which enabled the

Table 2 | Summary statistics for *M. genitalium*

Sequencing summary	
Number of instrument runs	1
Size of fibre-optic slide	60 × 60 mm ²
Run time/number of cycles	243 min/42
High quality reads	306,178
Average read length (bases)	110
Number of bases in high quality reads	33,655,553
Bases with a Phred score of 20 and above	26,753,540
Re-sequencing	
Reads mapped to single locations	238,066
Number of bases in mapped reads	27,687,747
Individual read insertion error rate	1.67%
Individual read deletion error rate	1.60%
Individual read substitution error rate	0.68%
Re-sequencing consensus	
Average over sampling	× 40
Coverage, all ($Z \geq 4$)	99.9% (98.2%)
Consensus accuracy, all ($Z \geq 4$)	99.97% (99.996%)
Consensus insertion error rate, all ($Z \geq 4$)	0.02% (0.003%)
Consensus deletion error rate, all ($Z \geq 4$)	0.01% (0.002%)
Consensus substitution error rate, all ($Z \geq 4$)	0.001% (0.0003%)
Number of contigs	10
<i>De novo</i> assembly	
Coverage, all ($Z \geq 4$)	96.54% (95.27%)
Consensus accuracy, all ($Z \geq 4$)	99.96% (99.994%)
Number of contigs	25
Average contig size (kb)	22.4

The individual read error rates are referenced to the total number of bases in mapped reads.

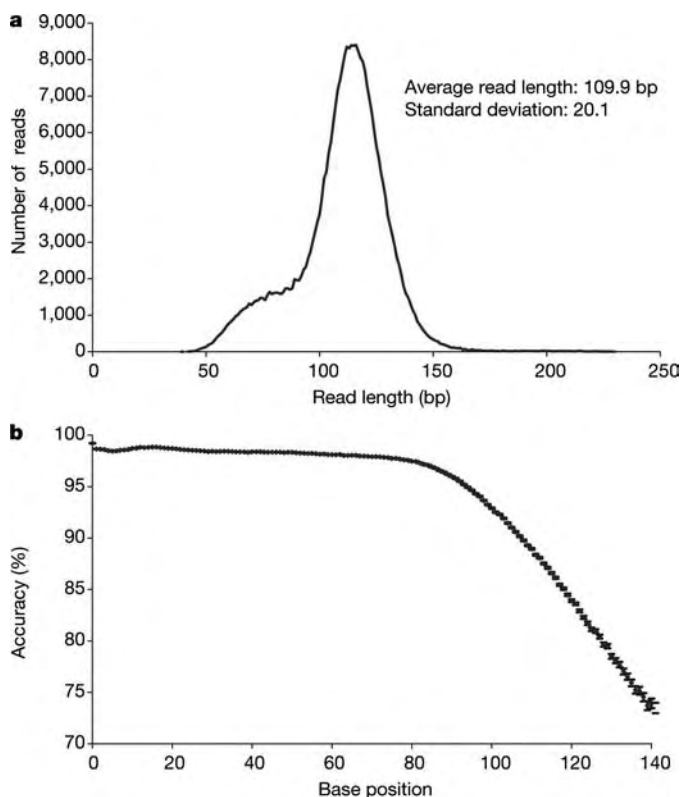


Figure 4 | *M. genitalium* data. a, Read length distribution for the 306,178 high-quality reads of the *M. genitalium* sequencing run. This distribution reflects the base composition of individual sequencing templates. **b**, Average read accuracy, at the single read level, as a function of base position for the 238,066 mapped reads of the same run.

prediction of significant improvements in the integrated circuit at the start of its development cycle¹⁸.

METHODS

Emulsion-based clonal amplification. The simultaneous amplification of fragments is achieved by isolating individual DNA-carrying beads in separate $\sim 100\text{-}\mu\text{m}$ aqueous droplets (on the order of $2 \times 10^6\text{ ml}^{-1}$) made through the creation of a PCR-reaction-mixture-in-oil emulsion. (Fig. 1b; see also Supplementary Methods). The droplets act as separate microreactors in which parallel DNA amplifications are performed, yielding approximately 10^7 copies of a template per bead; $800\text{ }\mu\text{l}$ of emulsion containing 1.5 million beads are prepared in a standard 2-ml tube. Each emulsion is aliquoted into eight PCR tubes for amplification. After PCR, the emulsion is broken to release the beads, which include beads with amplified, immobilized DNA template and empty beads (Supplementary Methods). We then enrich for template-carrying beads (Supplementary Methods). Typically, about 30% of the beads will have DNA, producing 450,000 template-carrying beads per emulsion reaction. The number of emulsions prepared depends on the size of the genome and the expected number of runs required to achieve adequate over sampling. The 580-kb *M. genitalium* genome, sequenced on one $60 \times 60\text{ mm}^2$ fibre-optic slide, required 1.6 ml of emulsion. A human genome, over sampled ten times, would require approximately 3,000 ml of emulsion.

Bead loading into picolitre wells. The enriched template-carrying beads are deposited by centrifugation into open wells (Fig. 1c), arranged along one face of a $60 \times 60\text{ mm}^2$ fibre-optic slide. The beads (diameter $\sim 28\text{ }\mu\text{m}$) are sized to ensure that no more than one bead fits in most wells (we observed that 2–5% of filled wells contain more than one bead). Loading 450,000 beads (from one emulsion preparation) onto each half of a $60 \times 60\text{ mm}^2$ plate was experimentally found to limit bead occupancy to approximately 35% of all wells, thereby reducing chemical and optical crosstalk between wells. A mixture of smaller beads that carry immobilized ATP sulphurylase and luciferase necessary to generate light from free pyrophosphate are also loaded into the wells to create the individual sequencing reactors (Supplementary Methods).

Image capture. A bead carrying 10 million copies of a template yields approximately 10,000 photons at the CCD sensor, per incorporated nucleotide. The generated light is transmitted through the base of the fibre-optic slide and detected by a large format CCD ($4,095 \times 4,096$ pixels). The images are processed to yield sequence information simultaneously for all wells containing template-carrying beads. The imaging system was designed to accommodate a large number of small wells and the large number of optical signals being generated from individual wells during each nucleotide flow. Once mounted, the fibre-optic slide's position does not shift; this makes it possible for the image analysis software to determine the location of each well (whether or not it contains a DNA-carrying bead), based on light generation during the flow of a pyrophosphate solution, which precedes each sequencing run. A single well is imaged by approximately nine $15\text{ }\mu\text{m}$ pixels. For each nucleotide flow, the light intensities collected by the pixels covering a particular well are summed to generate a signal for that particular well at that particular nucleotide flow. Each image captured by the CCD produces 32 megabytes of data. In order to perform all of the necessary signal processing in real time, the control computer is fitted with an accessory board (Supplementary Methods), hosting a 6 million gate Field Programmable Gate Array (FPGA)^{19,20}.

De novo shotgun sequence assembler. A *de novo* flow-space assembler was developed to capture all of the information contained in the original flow-based signal trace. It also addresses the fact that existing assemblers are not optimized for 80–120-bases reads, particularly with respect to memory management due to the increased number of sequencing reads needed to achieve equivalent genome coverage. (A completely random genome covered with 100-bases reads requires approximately 50% more reads to yield the same number of contiguous regions (contigs) as achieved with 700-bases reads, assuming the need for a 30-bases overlap between reads²¹.) This assembler consists of a series of modules: the Overlapper, which finds and creates overlaps between reads; the Unitigger, which constructs larger contigs of overlapping sequence reads; and the Multialigner, which generates consensus calls and quality scores for the bases within each contig (Supplementary Methods). (The names of the software modules are based on those performing related functions in other assemblers developed previously²².)

Received 6 May; accepted 10 June 2005.

Published online 31 July 2005.

1. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**, 5463–5467 (1977).
2. Prober, J. M. et al. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238**, 336–341 (1987).
3. NIH News Release. NHGRI seeks next generation of sequencing technologies. 14 October 2004 (<http://www.genome.gov/12513210>).
4. Nyren, P., Pettersson, B. & Uhlen, M. Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Anal. Biochem.* **208**, 171–175 (1993).
5. Ronaghi, M. et al. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* **242**, 84–89 (1996).
6. Jacobson, K. B. et al. Applications of mass spectrometry to DNA sequencing. *GATA* **8**, 223–229 (1991).
7. Bains, W. & Smith, G. C. A novel method for nucleic acid sequence determination. *J. Theor. Biol.* **135**, 303–307 (1988).
8. Jett, J. H. et al. High-speed DNA sequencing: an approach based upon fluorescence detection of single molecules. *Biomol. Struct. Dynam.* **7**, 301–309 (1989).
9. Tawfik, D. S. & Griffiths, A. D. Man-made cell-like compartments for molecular evolution. *Nature Biotechnol.* **16**, 652–656 (1998).
10. Ghadessy, F. J., Ong, J. L. & Holliger, P. Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl Acad. Sci. USA* **98**, 4552–4557 (2001).
11. Dressman, D., Yan, H., Traverso, G., Kinzler, K. W. & Vogelstein, B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl Acad. Sci. USA* **100**, 8817–8822 (2003).
12. Ronaghi, M., Uhlen, M. & Nyren, P. A sequencing method based on real-time pyrophosphate. *Science* **281**, 363–365 (1998).
13. Fraser, C. M. et al. The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403 (1995).
14. Tettelin, H. et al. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**, 498–506 (2001).
15. Leamon, J. H. et al. A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions. *Electrophoresis* **24**, 3769–3777 (2003).
16. Ronaghi, M. Pyrosequencing sheds light on DNA sequencing. *Genome Res.* **11**, 3–11 (2001).
17. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
18. Moore, G. E. Cramming more components onto integrated circuits. *Electronics* **38**(8) (1965).
19. Mehta, K., Rajesh, V. A. & Veeraswamy, S. FPGA implementation of VXIbus interface hardware. *Biomed. Sci. Instrum.* **29**, 507–513 (1993).
20. Fagin, B., Watt, J. G. & Gross, R. A special-purpose processor for gene sequence analysis. *Comput. Appl. Biosci.* **9**, 221–226 (1993).
21. Lander, E. S. & Waterman, M. S. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231–239 (1988).
22. Myers, E. W. Toward simplifying and accurately formulating fragment assembly. *J. Comput. Biol.* **2**, 275–290 (1995).
23. Ogawa, T. et al. Increased Productivity For Core Labs Using One Polymer and One Array Length for Multiple Applications. Poster P108-T. ABRF '05: Biomolecular Technologies: Discovery to Hypotheses (Savannah, Georgia, 5–8 February 2005); also available as Applied Biosystems 3730x/DNA Analyzer Specification Sheet (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge P. Dacey and the support of the Operations groups of 454 Life Sciences. This research was supported in part by the US Department of Health and Human Services under NIH grants.

Author Information Sequences for *M. genitalium* and *S. pneumoniae* are deposited at DDBJ/EMBL/GenBank under accession numbers AAGX01000000 and AAGY01000000, respectively. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to J.M.R. (jrothberg@454.com).

Discovery of a bright quasar without a massive host galaxy

Pierre Magain¹, G  r  ldine Letawe¹, Fr  d  ric Courbin², Pascale Jablonka^{2,3,4}, Knud Jahnke⁵, Georges Meylan² & Lutz Wisotzki⁵

A quasar is thought to be powered by the infall of matter onto a supermassive black hole at the centre of a massive galaxy^{1,2}. Because the optical luminosity of quasars exceeds that of their host galaxy, disentangling the two components can be difficult. This led in the 1990s to the controversial claim of the discovery of 'naked' quasars^{3–7}. Since then, the connection between quasars and galaxies has been well established⁸. Here we report the discovery of a quasar lying at the edge of a gas cloud, whose size is comparable to that of a small galaxy, but whose spectrum shows no evidence for stars. The gas in the cloud is excited by the quasar itself. If a host galaxy is present, it is at least six times fainter than would normally be expected^{8,9} for such a bright quasar. The quasar is interacting dynamically with a neighbouring galaxy, whose gas might be feeding the black hole.

HE0450–2958 is a bright quasar ($M_V = -25.8$) at a redshift of $z = 0.285$, associated with powerful infrared emission^{10,11}. Early imaging revealed a quasar–galaxy pair with signs of violent dynamical interaction¹², and a starburst in the galaxy. A collision between the two systems about 10⁸ years ago probably triggered both the starburst and the quasar activity¹³.

We use the MCS (Magain–Courbin–Sohy) deconvolution technique^{14–17} on our new Hubble Space Telescope (HST) images to explore the vicinity of the quasar (Fig. 1). These images confirm the strongly irregular shape of the companion galaxy, indicative of gravitational interaction with the quasar. However, the most interesting finding is that we found no significant host galaxy centred on the quasar position—we quantify this statement below. The most prominent feature is a diffuse, albeit compact, component that is just beside the quasar (and which we call 'the blob'), with a possible very faint extension around the quasar. If we interpreted this as an image of the host galaxy, we would have the surprising result that the quasar does not reside inside its host, but just next to it. Moreover, with a diameter of no more than 2,500 parsec (pc), the blob is comparable in size to M32, the dwarf companion of the Andromeda galaxy M31, but is about 200 times brighter than M32, with an absolute magnitude of $M_V = -22$.

Figure 2 shows a portion of the spatially deconvolved two-dimensional optical spectrum. The full one-dimensional spectra of the quasar, blob and companion galaxy are displayed on Fig. 3. Although the spectrum of the galaxy shows a strong stellar continuum, with low excitation emission lines typical of gas in star-forming regions, the spectrum of the blob does not show any hint of a continuum. Rather, it consists of a series of emission lines whose intensity ratios indicate excitation by a much harder spectrum than a stellar one. The absence of a continuum component yields the second surprising result: where we would expect stellar light from a host galaxy, we see

only an off-centre structure consisting essentially of gas excited by the quasar radiation.

This non-detection of a significant host galaxy around HE0450–2958 may mean that either it is too compact to be separated from the quasar, or that its surface brightness is below our detection limit. These possibilities are investigated by adding to the HST images an artificial elliptical galaxy centred on the quasar position and properly convolved with the HST point spread function (PSF). If we adopt an absolute magnitude of $M_V = -23$, corresponding to the typical host galaxies expected for such a quasar luminosity⁸, and if we decrease its size until it can no longer be detected in the deconvolved image, we obtain an upper limit of 100 pc for the half-light radius $R_{1/2}$. This is much too small for a quasar host galaxy, for which $R_{1/2}$ typically ranges between 2,000 and 15,000 pc (ref. 8).

To set an upper limit on the surface brightness of the putative host, we fix $R_{1/2}$ to 10 kpc, well representative of host galaxies of luminous quasars⁸, and we progressively decrease the luminosity until it falls below our detection limit. This gives an upper limit for the stellar population around the quasar; we discuss this limit below. The absence of any detectable extended structure is confirmed by applying standard PSF subtraction techniques. The sensitivity of our results to the presence of a host galaxy is illustrated by Fig. 4, which shows that, if a typical host galaxy were present around the quasar, we would indeed detect it.

The mass of the central black hole of a quasar of absolute magnitude $M_V = -25.8$ is predicted to be about 8×10^8 solar masses for an accretion at half the Eddington limit with a 10% radiative efficiency. This value is in fair agreement with the estimate derived from the width of the H β line¹⁸. Using the scaling relations recently established between the black-hole mass and the host-galaxy spheroid luminosity¹⁹, we can compare the predicted magnitude of the host galaxy with our upper limit. For this purpose, we have to make assumptions about the stellar population in the host galaxy. We consider two extreme cases.

On the one hand, we assume a smooth spheroid of old stars (ten billion years). The predicted magnitude²⁰ of the host is then $M_V = -23.0$ and our upper limit $M_V = -21.2$, that is, five times fainter than expected. On the other hand, in the case of young stars, of an age similar to the starburst population in the companion galaxy (130 million years), the predicted magnitude amounts to $M_V = -23.5$ and our upper limit becomes $M_V = -20.5$, that is, 16 times fainter. Any reasonable estimate should lie between these two extremes.

As a complementary approach, avoiding uncertainties related to the black-hole mass estimates, we compare our upper limits directly to the magnitudes measured for typical quasar host galaxies. We use the HST sample of 17 quasar host galaxies

¹Institut d'Astrophysique et de G  ophysique, Universit   de Li  ge, All  e du 6 Ao  t, 17, B  t B5C, B-4000 Li  ge, Belgium. ²Laboratoire d'Astrophysique, Ecole Polytechnique F  d  rale de Lausanne (EPFL), Observatoire, CH-1290 Sauvigny, Switzerland. ³Observatoire de l'Universit   de Gen  ve, CH-1290 Sauvigny, Switzerland. ⁴GEPI, UMR 8111, Observatoire de Paris, France. ⁵Astrophysikalisches Institut Potsdam, An der Sternwarte 16, D-14482 Potsdam, Germany.

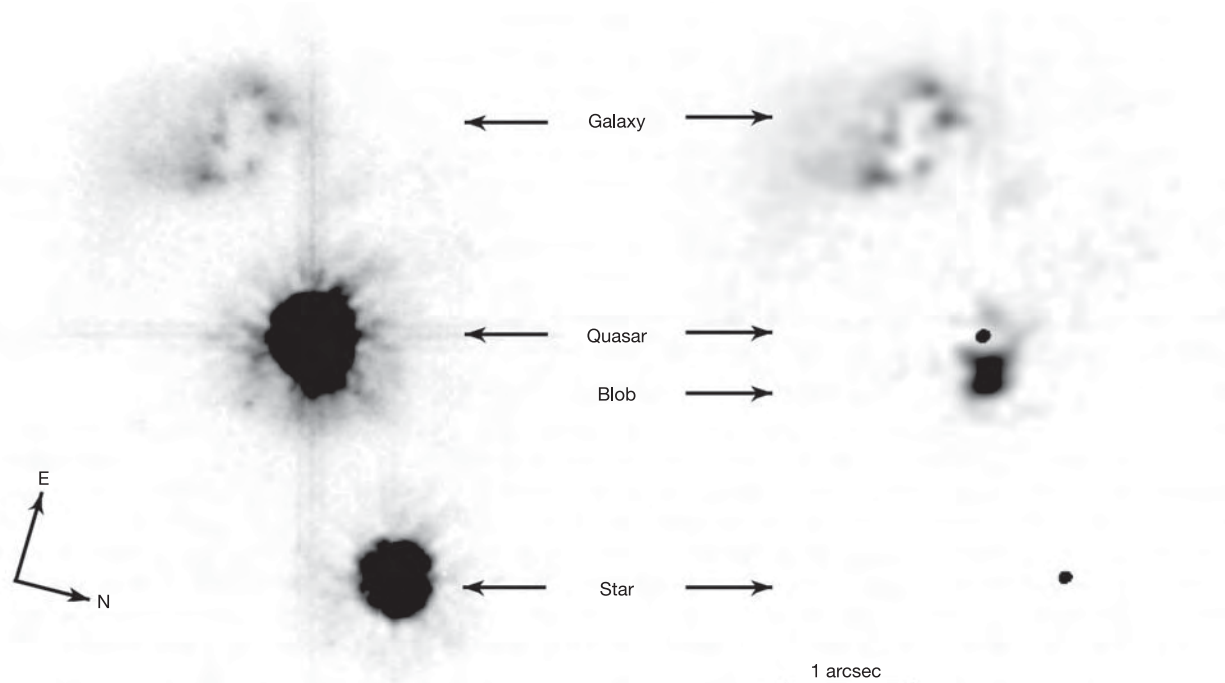


Figure 1 | HST images of the quasar and its immediate surroundings. The original HST image of HE0450–2958 is shown on the left, and the result of the deconvolution with the MCS algorithm displayed on the right. The images were obtained on 1 October 2004 with the High Resolution Channel (HRC) of the Advanced Camera for Surveys (ACS) onboard HST. Six dithered exposures of the quasar field were taken through the F606W filter, three short ones (30 s) and three longer ones (330 s). In contrast with previous HST observations of quasar host galaxies, a significant fraction of the observing time was devoted to the PSF characterization, by observing, during the same orbit, both the nearest bright star and the quasar, always placing the star at the same location as the quasar on the detector. This

observational strategy ensures that the PSF temporal and spatial variations are minimized. Because of the extended wings of the PSF, the blob of gas just beside the quasar can only be seen after careful processing of the images. No other related feature is found in the vicinity of the quasar–galaxy pair. From the observed fluxes, and applying the appropriate distance and spectral corrections, we obtain absolute magnitudes of $M_V = -22$ for the blob and $M_V = -23$ for the companion galaxy. After correction for dust extinction, the latter amounts to $M_V = -26$. All magnitudes mentioned in the text have been similarly corrected, and computed with $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

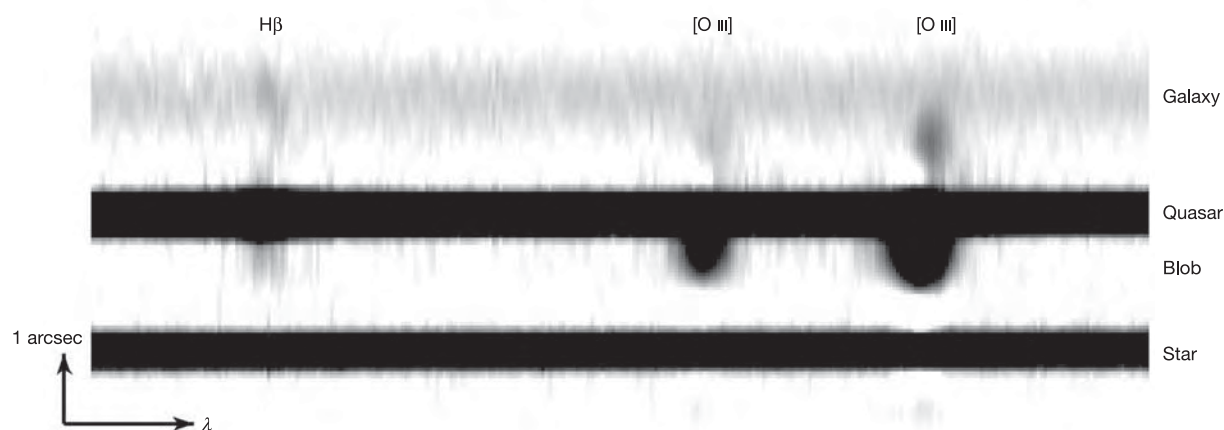


Figure 2 | Two-dimensional VLT spectra of the quasar and neighbouring objects. These long-slit spectra have been deconvolved in the spatial direction, in order to separate the different objects. The spectral direction is horizontal, while the spatial direction is vertical. The emission lines just below the quasar spectrum are, from left to right, the hydrogen H β (486.1 nm) and the oxygen [O III] (495.9 nm/500.7 nm) doublet corresponding to the blob of ionized gas. Weaker emission lines are also detected in between the quasar and the companion galaxy, showing that some ionized gas is also present there. The vertical arrow gives the spatial

scale (1"). Our spectroscopic observations of HE0450–2958 were carried out in a new way, taking advantage of the multi-slit mode of the VLT/FORS1 to record simultaneously the spectra of the quasar and of its host galaxy, as well as the spectra of nearby isolated stars needed for subsequent determination of the PSF. The main slit covers not only the quasar, but also the companion galaxy and a nearby object identified as a foreground star, which is used together with the isolated stars for an accurate determination of the PSF.

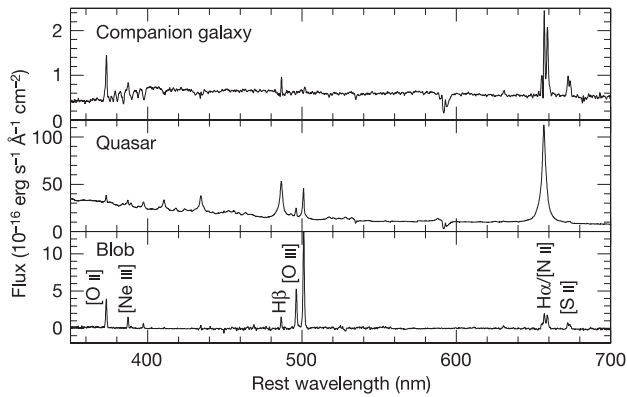


Figure 3 | Spectra of the companion galaxy, quasar, and blob. Top, integrated one-dimensional spectrum of the companion galaxy. Middle, spectrum of the quasar. Bottom, spectrum of the blob. Note the absence of continuum light in the blob, which would be due to stars, and the emission line ratios (for example, $[O III]/H\beta$) indicating ionization by the quasar radiation. The Balmer decrement ($H\alpha/H\beta$ intensity ratio) indicates that, contrary to the quasar and blob, the companion galaxy is heavily reddened by dust. It is therefore the companion galaxy, and not the quasar, that is the strong infrared emitter detected by the IRAS satellite¹⁰. This also indicates that absorption of light by dust cannot explain the non-detection of a host galaxy around the quasar.

from ref. 8. After conversion of their values to our cosmology ($H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$), these quasars have absolute magnitudes in the range $-26.8 < M_V < -23.2$, with an average of $M_V = -24.3$, somewhat fainter than the present quasar ($M_V = -25.8$). A straight-line fitting to the host-magnitude versus the quasar-magnitude relation gives an expected host magnitude of $M_V = -23.2$ for a quasar of $M_V = -25.8$, and the scatter around the mean relation is $\sigma = 0.54$ (Fig. 5). Note that this value of $M_V = -23.2$ falls nicely between the two extreme cases considered in the previous paragraph. Our upper limit is between 6 and 12 times fainter than the expected value, depending on the assumed stellar age. In the most conservative case (that is, a smooth spheroidal distribution of old stars), the

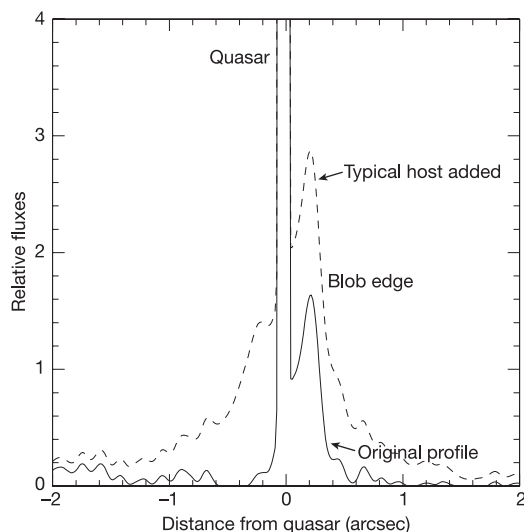


Figure 4 | Intensity profile through the quasar position. The solid line is the N-S profile of the deconvolved image. The peak just on the right of the quasar corresponds to the edge of the blob. The dashed line is the corresponding profile of the deconvolved image after an elliptical host (de Vaucouleurs profile) with $R_{1/2} = 10 \text{ kpc}$ and $M_V = -23$ has been artificially added to the original images. The comparison of the two profiles shows that any host galaxy with the expected luminosity would be easily detected.

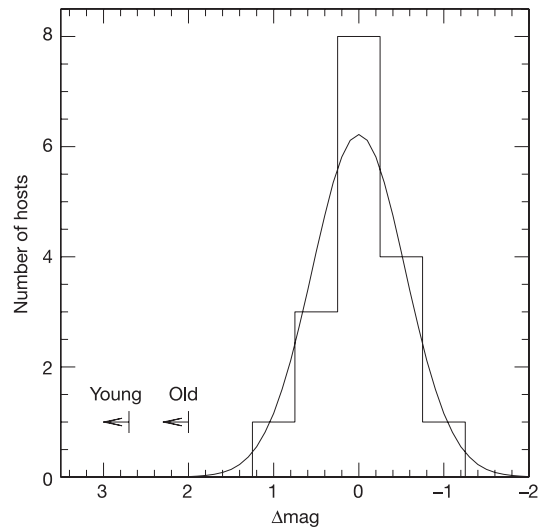


Figure 5 | Upper limits on the host magnitudes. The histogram gives the dispersion of the host magnitudes around the mean trend in the HST luminous quasars sample⁸. The curve represents a gaussian fit to these data. Our upper limits for a smooth spheroidal distribution of either old or young stars are indicated. They both deviate from the mean by a large amount of 3.7 and 5σ , respectively.

deviation from the average relation is 3.7σ , which is highly significant. Moreover, we should point out that such a conservative case is not really expected for galaxies involved in collisions, which generally display a distorted geometry (for example, tidal tails) and active star formation. Taking these effects into account would make the deviation even more significant.

Finally, the hydrogen emission line ratios, which indicate a strong reddening of the companion galaxy, are close to the theoretically expected values in the quasar and blob. This suggests that obscuration by dust is very limited in the vicinity of the quasar and cannot explain the non-detection of the host galaxy.

This peculiar quasar, therefore, has a host galaxy (if any) that is significantly less luminous than expected from its nuclear luminosity and black hole mass. Moreover, it suffers a strong dynamical interaction with an ultra-luminous infrared galaxy (a rare class of galaxies, systematically involved in collisions¹³). These two peculiarities are most probably related. One might suggest that the host galaxy has disappeared from our view as a result of the collision, but it is hard to imagine how the complete disruption of a galaxy could happen. An alternative suggestion would be that an isolated black hole may have captured gas and become a quasar while crossing the disk of the neighbouring galaxy with a low relative speed (the radial velocity difference between the quasar and different parts of the galaxy ranges between -60 and $+200 \text{ km s}^{-1}$). However, such a gravitational accretion of matter is very inefficient and the dynamical interaction of a large galaxy with a black hole of 8×10^8 solar masses would probably not induce perturbations as strong as those observed in the companion galaxy.

Another possibility would be that the black hole of HE0450-2958 lies in a galaxy with not only a stellar content much lower than average, but also with an important dark halo, a "dark galaxy"²¹. The interaction with such a massive object could more easily explain the peculiarities of the neighbouring galaxy as well as the capture of gas, resulting in the ignition of the quasar.

Received 23 March; accepted 29 June 2005.

1. Richstone, D. et al. Supermassive black holes and the evolution of galaxies. *Nature* **395**, A14-A18 (1998).
2. Dunlop, J. S. et al. Quasars, their host galaxies and their central black holes. *Mon. Not. R. Astron. Soc.* **340**, 1095-1135 (2003).

3. Bahcall, J. N., Kirhakos, S. & Schneider, D. P. HST images of nearby luminous quasars. *Astrophys. J.* **435**, L11–L14 (1994).
4. Boyce, P. J., Disney, M. J. & Bleaken, D. G. Quasar host galaxy images from the Hubble Space Telescope archive. *Mon. Not. R. Astron. Soc.* **302**, L39–L44 (1999).
5. Bahcall, J. N., Kirhakos, S., Saxe, D. H. & Schneider, D. P. Hubble Space Telescope images of a sample of 20 nearby luminous quasars. *Astrophys. J.* **479**, 642–658 (1997).
6. Disney, M. J. *et al.* Interacting elliptical galaxies as hosts of intermediate redshift quasars. *Nature* **376**, 150–153 (1995).
7. Hutchings, J. B. Quasars—hosts of possibilities. *Nature* **376**, 118–119 (1995).
8. Floyd, D. J. E. *et al.* The host galaxies of luminous quasars. *Mon. Not. R. Astron. Soc.* **355**, 196–220 (2004).
9. McLure, R. J. *et al.* A comparative HST imaging study of the host galaxies of radio-quiet quasars, radio-loud quasars and radio galaxies—I. *Mon. Not. R. Astron. Soc.* **308**, 377–404 (1999).
10. de Grijp, M. H. K., Lub, J. & Miley, G. K. Warm IRAS sources. I. A catalogue of AGN candidates from the point source catalog. *Astron. Astrophys. Suppl.* **70**, 95–114 (1987).
11. Low, F. J., Cutri, R. M., Huchra, J. P. & Kleinman, S. G. Infrared colour-selected quasars and Seyfert 1 galaxies. *Astrophys. J.* **327**, L41–L45 (1988).
12. Boyce, P. J. *et al.* The host galaxies of IRAS-selected quasi-stellar objects. *Astrophys. J.* **473**, 760–762 (1996).
13. Canalizo, G. & Stockton, A. Quasi-stellar objects, ultraluminous infrared galaxies, and mergers. *Astrophys. J.* **555**, 719–743 (2001).
14. Magain, P., Courbin, F. & Sohy, S. Deconvolution with correct sampling. *Astrophys. J.* **494**, 472–477 (1998).
15. Courbin, F., Magain, P., Kirkove, M. & Sohy, S. A method for spatial deconvolution of spectra. *Astrophys. J.* **529**, 1136–1144 (2000).
16. Courbin, F. *et al.* On-axis spatially resolved spectroscopy of low redshift quasar host galaxies: HE 1503 + 0228, at $z = 0.135$. *Astron. Astrophys.* **394**, 863–872 (2002).
17. Letawe, G. *et al.* On-axis spectroscopy of the $z = 0.144$ radio-loud quasar HE 1434–1600: an elliptical host with a highly ionized ISM. *Astron. Astrophys.* **424**, 455–464 (2004).
18. Kaspi, S. *et al.* Reverberation measurements for 17 quasars and the size–mass–luminosity relations in active galactic nuclei. *Astrophys. J.* **533**, 631–649 (2000).
19. McLure, R. J. & Dunlop, J. S. On the black hole–bulge mass relation in active and inactive galaxies. *Mon. Not. R. Astron. Soc.* **331**, 795–804 (2002).
20. Bruzual, G. & Charlot, S. Stellar population synthesis at the resolution of 2003. *Mon. Not. R. Astron. Soc.* **344**, 1000–1028 (2003).
21. Ferrarese, L. Beyond the bulge: A fundamental relation between supermassive black holes and dark matter halos. *Astrophys. J.* **578**, 90–97 (2002).

Acknowledgements P.J. is on leave from GEPI. This work has been supported by the PPS Science Policy (Belgium), by PRODEX (ESA), and by the Swiss National Science Foundation. The observations were obtained with the ESO/VLT (Paranal, Chile) and with the NASA/ESA Hubble Space Telescope.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.M. (Pierre.Magain@ulg.ac.be).

Extreme oxygen isotope ratios in the early Solar System

Jérôme Aléon^{1†}, François Robert², Jean Duprat³ & Sylvie Derenne⁴

The origins of the building blocks of the Solar System can be studied using the isotopic composition of early planetary and meteoritic material. Oxygen isotopes in planetary materials show variations at the per cent level that are not related to the mass of the isotopes^{1,2}; rather, they result from the mixture of components having different nucleosynthetic or chemical origins^{1–3}. Isotopic variations reaching orders of magnitude in minute meteoritic grains are usually attributed to stellar nucleosynthesis before the birth of the Solar System, whereby different grains were contributed by different stars^{4,5}. Here we report the discovery of abundant silica-rich grains embedded in meteoritic organic matter, having the most extreme $^{18}\text{O}/^{16}\text{O}$ and $^{17}\text{O}/^{16}\text{O}$ ratios observed (both $\sim 10^{-1}$) together with a solar silicon isotopic composition. Both O and Si isotopes indicate a single nucleosynthetic process. These compositions can be accounted for by one of two processes: a single exotic evolved star seeding the young Solar System⁶, or irradiation of the circumsolar gas by high energy particles accelerated during an active phase of the young Sun. We favour the latter interpretation, because the observed compositions are usually not expected from nucleosynthetic processes in evolved stars, whereas they are predicted by the selective trapping of irradiation products.

During ion imaging study of acid insoluble organic matter (IOM) isolated from the Murchison meteorite⁷, we discovered a surprising ^{18}O hotspot in one image. After systematic oxygen and silicon isotopic mapping using the IMS 1270 ion microprobe at CRPG-Nancy, we found 31 silicon-rich grains with extreme ^{17}O and ^{18}O excesses embedded within the IOM (Fig. 1). Details of sample handling and analytical procedures are provided as Supplementary Information, along with the complete isotopic data set. Oxygen isotope ratios define a mixing line between the solar and the most extreme values (Fig. 2a). This line could result either from mixing, within the ion probe spot, of O from the grains and from IOM in various proportions, or from the indigenous isotopic composition of the grains. Whatever the exact amplitude of these two effects, the data demonstrate that the mixing occurs with a single and extremely heavy oxygen reservoir (with $^{17}\text{O}/^{16}\text{O} \geq 7.7 \times 10^{-2}$, $^{18}\text{O}/^{16}\text{O} \geq 1.2 \times 10^{-1}$ and $^{18}\text{O}/^{17}\text{O} \leq 1.5$, Table 1).

Silicon isotopic data are almost indistinguishable from the terrestrial and meteoritic reservoirs. The weighted means given as per mil deviations from the terrestrial reference values (termed $\delta^{29}\text{Si}$ and $\delta^{30}\text{Si}$ for the $^{29}\text{Si}/^{28}\text{Si}$ and $^{30}\text{Si}/^{28}\text{Si}$ ratios, respectively) are $\delta^{29}\text{Si} = 9.7 \pm 2.2\text{‰}$ and $\delta^{30}\text{Si} = 7.4 \pm 3.0\text{‰}$ (95% confidence interval). Further examination by scanning electron microscopy (SEM) and energy dispersive X-ray spectroscopy revealed that these grains are only constituted of Si and O (Fig. 3a, b). Although their exact crystallinity and stoichiometry cannot be determined

without ambiguity, several grains are platy and faceted (Fig. 3a) suggesting they could be silica (they are hereafter referred to as 'silica-rich'). Coated with IOM, the surfaces of the grains were not detectable by SEM and became visible only after IOM destruction upon laser treatment (Fig. 3c, d). Their abundance, based on their relative surface area and the yield of IOM extraction, ranges from 250 p.p.b. to 1.8 p.p.m. (by weight), with a best estimate of 870 p.p.b. This abundance is comparable to that of presolar oxide grains⁸ but lower than that of presolar silicates in other primitive meteorites by about an order of magnitude^{9,10}.

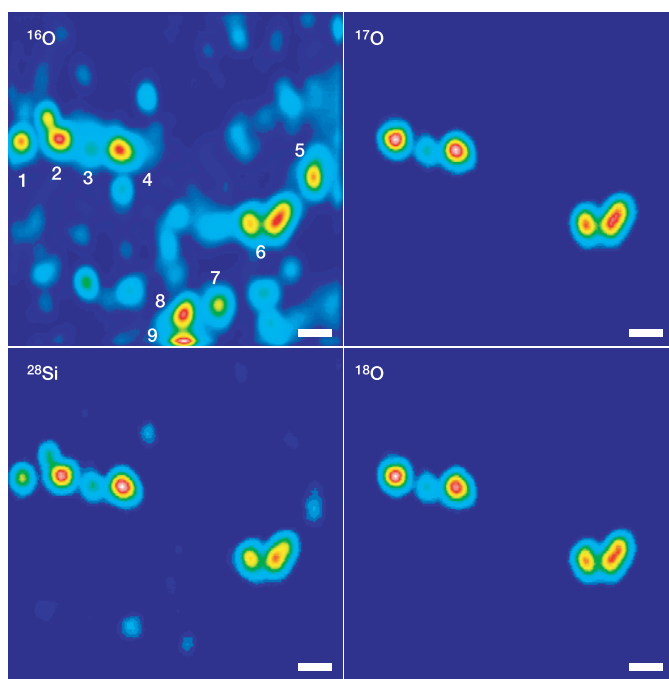


Figure 1 | Ion images of ^{16}O , ^{17}O , ^{18}O and ^{28}Si in organic pellets. Images were smoothed with a 7×7 pixels kernel representing the convolution by the ion probe beam. Intensities are normalized to show comparable variations, and increase linearly from dark blue to white. Grains labelled 1 to 9 were examined by field-emission SEM. Grains 2, 3, 4 and cluster 6 are silica-rich grains with oxygen isotope anomalies (M3, M5, M4 and M6 in Table 1, respectively), grain 1 is an Al-rich silicate with solar isotopic composition, grains 5 and 8 are solar corundum grains, grain 7 is a solar chromite and grain 9 is a big solar spinel cut by the edge of the image. ^{17}O and ^{18}O intensities of solar grains are so low compared to the anomalous grains that they have been wiped out in smoothed images. Scale bars, $2\text{ }\mu\text{m}$.

¹Centre de Recherches Pétrographiques et Géochimiques, 15 rue Notre Dame des Pauvres, BP20, 54501 Vandoeuvre-les-Nancy, France. ²Laboratoire d'Etude de la Matière Extraterrestre, Muséum National d'Histoire Naturelle, 61 rue Buffon, 75005 Paris, France. ³Centre de Spectrométrie Nucléaire et de Spectrométrie de Masse, Bat 104, 91405 Orsay Campus, France. ⁴Laboratoire de Chimie Bioorganique et Organique Physique, Ecole Nationale Supérieure de Chimie de Paris, 11 rue Pierre et Marie Curie, 75231 Paris Cedex 05, France. [†]Present address: Lawrence Livermore National Laboratory, Glenn T. Seaborg Institute, P.O. Box 808, L-231, Livermore, California 94550, USA.

Large O isotopic anomalies in meteoritic grains are usually attributed to stellar nucleosynthesis. However, the association of large ^{17}O and ^{18}O excesses together with solar Si isotope composition represents a challenge to stellar models. Extreme ratios reaching 10^{-1} have only been observed once¹¹—in HR 4049, a post-asymptotic giant branch (AGB) star entering the planetary nebula phase (Fig. 2b)—but no nucleosynthesis model of AGB stars could explain these compositions¹¹. Smaller ^{17}O and ^{18}O excesses have been reported in HD 101013, a Ba star¹² (Fig. 2b) that is an evolved star enriched in products of s-process nucleosynthesis. Both HD 101013 and HR 4049 are binaries, suggesting a mixing of an extreme O reservoir (observed in HR 4049) with a close to solar O reservoir, which can be the companion star¹¹. Our data indicate a single

endmember in contrast with the large isotopic variations observed in most populations of presolar grains^{4,5,8–10,13} and expected from multiple stellar sources and/or stellar condensation locations. This single source must account for almost 1 p.p.m. of the host meteorite, an abundance comparable to that of the total population of presolar oxide grains⁸.

Therefore, a first hypothesis to explain the composition of our grains would be the seeding of the young Solar System by the ejecta of a single star comparable to HR4049 and HD101013. In that respect, AGB stars have already been proposed to be a source, via s-process nucleosynthesis, of several extinct radionuclide excesses reported in meteoritic materials⁶. In a second hypothesis, where the observed heavy oxygen isotopes would be produced in several stars, a single nucleosynthesis process is required to produce both extreme ^{17}O and ^{18}O enrichments in evolved stars such as HR 4049, HD 101013 and the sources of the silica-rich grains. In the absence of Si isotopic anomalies, this process cannot be explosive hydrogen-burning during novae outbursts¹⁴ involving an evolved star and its companion star. The ^{18}O excesses rarely observed in evolved stars have been attributed to the presence in the stellar photosphere of shell helium-burning products with a fine-tuning of the temperature to preclude subsequent conversion of ^{18}O to ^{22}Ne (refs 12, 15). However, the mixing, occurring in different stars in variable proportions, of a ^{18}O -rich shell He-burning reservoir with an ^{17}O -rich envelope is expected to produce an O endmember that should exhibit large variations in the $^{18}\text{O}/^{17}\text{O}$ ratio, whereas the present grains are characterized by a constant $^{18}\text{O}/^{17}\text{O} = 1.65 \pm 0.42$ (2σ). Finally, if these grains are presolar, they probably acquired their intricate coating of IOM in the parent molecular cloud of the Solar System.

An alternative hypothesis would be the production of these anomalies within the Solar System itself. Astronomical observations of young stellar objects (YSOs) show that nascent solar-type stars undergo a period of intense X-ray activity, during which their circumstellar material could have endured bombardment by light charged particles with as much as 10^5 times the average fluence observed from the modern Sun¹⁶. Within the MeV energy range, the nuclear production of oxygen atoms is dominated by reactions on the most abundant neighbouring targets (N, O and Ne) that are in the gas phase. The irradiation of a solar gas¹⁷ by particles with

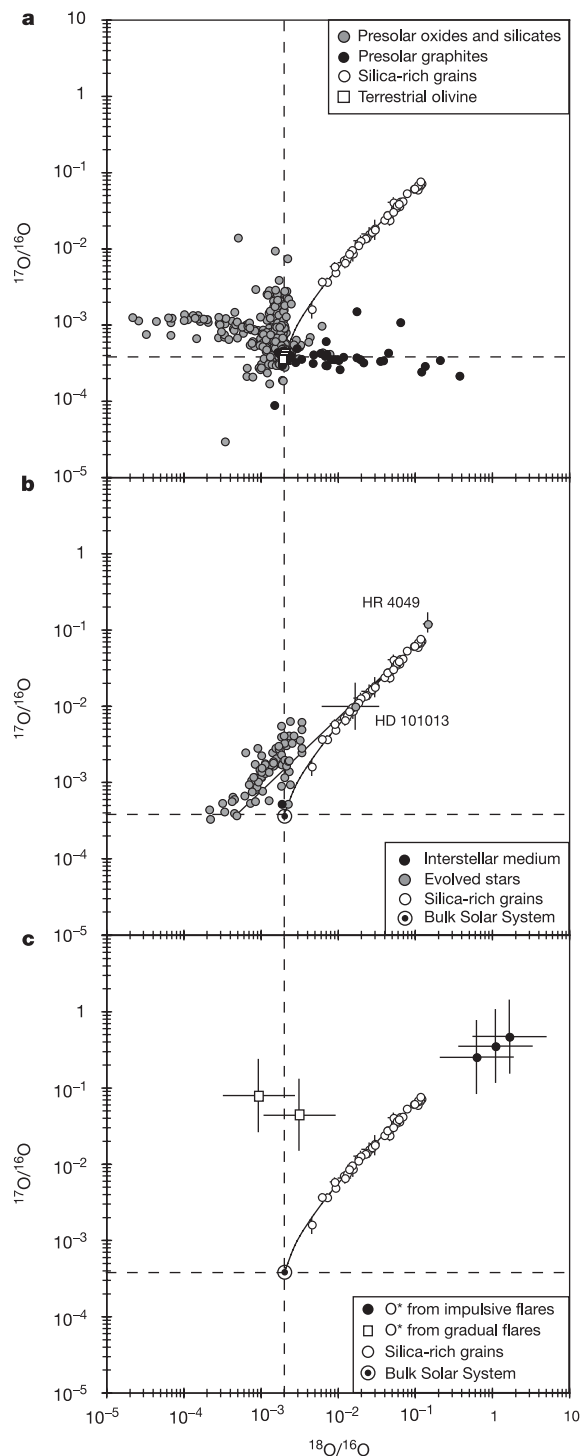


Figure 2 | Oxygen isotopic composition of anomalous silica-rich grains compared with presolar grains, evolved stars and irradiation models. The grains are shown by unshaded open circles on all panels. Error bars (2σ analytical errors) are mostly smaller than the symbol size. Dashed lines are terrestrial ratios in all panels. **a**, Comparison with presolar stardust grains found in meteorites (grey filled circles are oxide and silicate grains^{4,5,8–10} and black filled circles are graphite grains¹³). Terrestrial olivine grains were measured as an internal reference. The solid line is a mixing line between the $^{17,18}\text{O}$ -rich silica endmember and the solar endmember. **b**, Comparison with evolved stars^{11,12,15,23–29}. With the exception of HR 4049 where circumstellar CO_2 is measured¹¹, other data are photospheric CO . Error bars are shown only for HD 101013 and HR 4049 for clarity. The two solid lines correspond to the mixing of the extreme $^{17,18}\text{O}$ -rich endmember with the Solar System on one hand and with the most ^{17}O -rich hypothetical evolved star that can still reproduce the observed trend on the other hand. Note that the lowest silica-rich grain value is a duplicate analysis with a significant proportion of solar oxygen from the adjacent IOM. The area in-between the two lines shows thus the composition of all possible close-to-solar endmembers. Interestingly, the value of the present-day local interstellar medium (black filled circle) is in this field³⁰. **c**, Oxygen isotopic compositions of the reservoir O^* produced by irradiation of the circumsolar gas (that is, from O, Ne and N) by particles (protons (p), ^4He and ^3He) with characteristics of solar flares. Taking incident particles with energy spectra as $dN/dE = KE^{-\gamma}$, where N is the number of particles, and E their energy, the different values correspond to $\gamma = 3$ (lower value), and 4 and 5 (higher value) for impulsive flares ($p/^4\text{He} = p/^3\text{He} = 0.1$) and $\gamma = 2$ (right value) and 3 (left value) for gradual flares ($p/^4\text{He} = 0.01$; $^3\text{He}/^4\text{He} = 5 \times 10^{-4}$). Error bars are two-sigma analytical errors (2 s.e.m.).

Table 1 | Isotopic composition of silica-rich grains and nuclear-induced oxygen

Source of oxygen	$^{17}\text{O}/^{16}\text{O}$ (10^{-2})	$^{18}\text{O}/^{16}\text{O}$ (10^{-1})	$^{18}\text{O}/^{17}\text{O}$	Irradiation models
Solar O reference	3.82×10^{-2}	2.01×10^{-2}	5.25	
Measured endmember grains				
M3	6.74 (0.27)	1.14 (0.03)	1.69 (0.08)	
M4	7.21 (0.22)	1.20 (0.05)	1.67 (0.08)	
M5	6.16 (0.30)	0.98 (0.06)	1.56 (0.12)	
M6	6.94 (0.11)	1.15 (0.02)	1.66 (0.04)	
M7	6.00 (0.35)	1.07 (0.08)	1.78 (0.17)	
M9	6.65 (0.26)	1.07 (0.05)	1.61 (0.10)	
M10	6.91 (0.27)	1.08 (0.02)	1.56 (0.07)	
M16	7.71 (0.19)	1.16 (0.04)	1.50 (0.06)	
Irradiation models				
Ne target	300–2,400	17–260	0.5–1	(IM type) $\gamma = 3\text{--}5$; $\alpha/p = {}^3\text{He}/p = 0.1$
O target	12–8	6–22	5–27	(IM type) $\gamma = 3\text{--}5$; $\alpha/p = {}^3\text{He}/p = 0.1$
N target	315–170	$(2\text{--}8) \times 10^{-2}$	$(1\text{--}3) \times 10^{-3}$	(IM type) $\gamma = 3\text{--}5$; $\alpha/p = {}^3\text{He}/p = 0.1$
Solar gas target	25–50	6–16	2.4–3.5	(IM type) $\gamma = 3\text{--}5$; $\alpha/p = {}^3\text{He}/p = 0.1$
Solar gas target	4.5–8	$(0.9\text{--}3) \times 10^{-2}$	$(1\text{--}7) \times 10^{-2}$	(GR type) $\gamma = 2\text{--}3$; $\alpha/p = 0.01$; ${}^3\text{He}/\alpha = 5 \times 10^{-4}$

In values for endmember grains, 2σ errors are given between parentheses. γ , Slope of the energy distribution of incident particles assuming a $E^{-\gamma}$ power law. GR type, similar to gradual flares; IM type, similar to impulsive flares.

characteristics of impulsive solar flares¹⁸ results in a nuclear-induced oxygen reservoir (hereafter O*) with both $^{17}\text{O}/^{16}\text{O}$ and $^{18}\text{O}/^{16}\text{O}$ ratios at least an order of magnitude above the most extreme measured ratios (Table 1, Fig. 2c, Supplementary Information). To prevent a complete dilution of the O* anomaly in the large amounts of free solar oxygen expected to be released in the irradiation zone, O* must be isolated from the irradiated gas before condensation. Being nuclear-induced, O* is characterized by high recoil energy (~ 1 MeV). After leaving the irradiation zone, it could be retrieved from the gas phase by a condensation reaction with SiO: $\text{SiO}_{(\text{gas})} + \text{O}^* \rightarrow \text{SiO}_{2(\text{solid})}$.

Large SiO emission is commonly observed in high velocity proto-stellar molecular outflows, with inferred SiO abundance up to 10^4 that in quiescent clouds¹⁹. Such energetic SiO-rich outflows could be

the location for the isolation of the energetic O* reservoir and the subsequent condensation of the observed silica-rich micrometre-sized grains. In such a chemical reaction,^{29,30} Si excesses in SiO_2 from nuclear-induced Si* are expected to be negligible because the intermediate species $\text{Si}^*\text{O}_{(\text{gas})}$ will be diluted in the major solar $\text{SiO}_{(\text{gas})}$ reservoir. The observed mixing line can be explained by a mixture of 0.1–10% of the inferred O* reservoir with solar SiO. Assuming that the proportion of O* in the irradiated reservoir is similar to that of O* in the bulk meteorite, the inferred total proton fluence is about 10^{17} protons cm^{-2} (having an energy $E_p > 10$ MeV). Such a fluence is two orders of magnitude lower than that required to produce irradiation-induced ^{10}Be in meteoritic refractory inclusions^{20,21}, but is comparable to that required to produce the excesses of spallogenic ^{21}Ne and ^{38}Ar reported in individual meteoritic grains, including those from the Murchison meteorite²². If formed in protosolar outflows, the silica-rich grains could have fallen onto the surface of the accretion disk, where their intricate organic polymer coating would be assembled owing to an interstellar-type chemistry.

Interestingly, the isotopic anomalies in HR 4049 are thought to reflect the abundances of ^{16}O , ^{17}O and ^{18}O in the stellar winds¹¹. They were observed in CO_2 that could be the trapping reservoir for an anomalous O* via the reaction $\text{CO} + \text{O}^* \rightarrow \text{CO}_2$ similar to that of SiO_2 formation. This surprising similarity may point towards a widespread selective chemical trapping of a nucleosynthetic reservoir occurring in stellar ejecta/outflows that could explain the isotopic composition of both CO_2 in HR 4049 and the present SiO_2 -rich grains. Finally, theories of Solar System formation are widely based on the isotopic composition of stellar, interstellar and early Solar System materials. Most stars are born in multiple stellar clusters, but whether the Sun was born in association with massive OB stars or in a cluster of low mass stars is still an open question. Here we report a new population of extraterrestrial material; deciphering whether this material formed around the young Sun by a process capable of changing drastically the composition of circumstellar materials or whether it formed in the ejecta of a nearby exotic low mass star should shed light on the birth of our star.

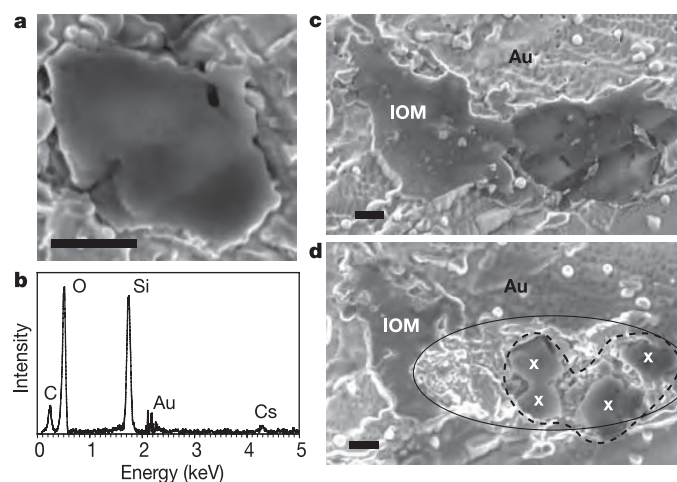


Figure 3 | Mineralogical observations of anomalous grains. **a**, Secondary electron image of euhedral silica-rich grain M3 having a platy and faceted morphology. **b**, Energy dispersive X-ray spectra of grain M3 after subtraction of the Au background. Prominent O and Si peaks suggest that M3 is silica-rich. C, Au and Cs peaks are due to the surrounding organic carbon, the residual of background subtraction and implantation of sputtering ions during ion probe analysis, respectively. **c**, IOM (insoluble organic matter) pellet containing the cluster M6 localized using a carbon image. **d**, Same as **c** after laser treatment. Residual IOM is indicated. The solid ellipse shows the impact area of the laser beam. Only IOM has been affected by the laser beam; even gold is intact. The dashed line encloses composite grain M6 (labelled 6 in Fig. 1), which appears to be a cluster of at least four anomalous silica-rich grains shown by crosses. Scale bars, 500 nm.

Received 26 January; accepted 17 June 2005.

1. Clayton, R. N., Grossman, L. & Mayeda, T. K. A component of primitive nuclear composition in carbonaceous meteorites. *Science* **182**, 485–487 (1973).
2. Clayton, R. N. Oxygen isotopes in meteorites. *Annu. Rev. Earth Planet. Sci.* **21**, 115–149 (1993).
3. McKeegan, K. D. M. & Leshin, L. A. in *Stable Isotope Geochemistry* (eds Valley, J. W. & Cole, D. R.) 279–318 (Mineralogical Society of America, Washington, 2001).

4. Nittler, L. R., Alexander, C. M. O'D., Gao, X., Walker, R. M. & Zinner, E. K. Interstellar oxide grains from the Tieschitz ordinary chondrite. *Nature* **370**, 443–446 (1994).
5. Clayton, D. D. & Nittler, L. R. Astrophysics with presolar stardust. *Annu. Rev. Astron. Astrophys.* **42**, 39–78 (2004).
6. Busso, M., Gallino, R. & Wasserburg, G. J. Short-lived nuclei in the early solar system: A low mass stellar source? *Publ. Astron. Soc. Austr.* **20**, 356–370 (2003).
7. Gardinier, A. *et al.* Solid state CP/MAS ^{13}C NMR of the insoluble organic matter of the Orgueil and Murchison meteorites: quantitative study. *Earth Planet. Sci. Lett.* **184**, 9–21 (2000).
8. Zinner, E. *et al.* Presolar spinel grains from the Murray and Murchison carbonaceous chondrites. *Geochim. Cosmochim. Acta* **67**, 5083–5095 (2003).
9. Nguyen, A. N. & Zinner, E. Discovery of ancient silicate stardust in a meteorite. *Science* **303**, 1496–1499 (2004).
10. Nagashima, K., Krot, A. N. & Yurimoto, H. Stardust silicates from primitive meteorites. *Nature* **428**, 921–924 (2004).
11. Cami, J. & Yamamura, I. Discovery of anomalous oxygen isotopic ratios in HR 4049. *Astron. Astrophys.* **367**, L1–L4 (2001).
12. Harris, M. J., Lambert, D. L. & Smith, V. V. Oxygen isotopic abundances in evolved stars. I. Six barium stars. *Astrophys. J.* **292**, 620–627 (1985).
13. Amari, S., Zinner, E. & Lewis, R. S. Large ^{18}O excesses in circumstellar graphite grains from the Murchison meteorite: Indication of a massive-star origin. *Astrophys. J.* **447**, L147–L150 (1995).
14. José, J., Hernanz, M., Amari, S., Lodders, K. & Zinner, E. The imprint of nova nucleosynthesis in presolar grains. *Astrophys. J.* **612**, 414–428 (2004).
15. Clayton, G. C. *et al.* An extremely large excess of ^{18}O in the hydrogen deficient carbon star HD 137613. *Astrophys. J.* **623**, L141–L144 (2005).
16. Feigelson, E. D., Garmire, G. P. & Pravdo, S. H. Magnetic flaring in the pre-main-sequence sun and implications for the early solar system. *Astrophys. J.* **572**, 335–349 (2002).
17. Anders, E. & Grevesse, N. Abundances of the elements: Meteoritic and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).
18. Reames, D. V. *et al.* Energy spectra of ions accelerated in impulsive and gradual solar events. *Astrophys. J.* **483**, 515–522 (1997).
19. Nisini, B., Codella, C., Giannini, T. & Richer, J. S. Observations of high-J SiO emission along the HH221 outflow. *Astron. Astrophys.* **395**, L25–L28 (2002).
20. McKeegan, K. D., Chaussidon, M. & Robert, F. Incorporation of short-lived ^{10}Be in a calcium-aluminum-rich inclusion from the Allende meteorite. *Science* **289**, 1334–1337 (2000).
21. Gounelle, M. *et al.* Extinct radioactivities and protosolar cosmic rays: self-shielding and light elements. *Astrophys. J.* **548**, 1051–1070 (2001).
22. Caffee, M. W., Hohenberg, C. M., Swindle, T. D. & Goswami, J. N. Evidence in meteorites for an active early sun. *Astrophys. J.* **313**, L31–L35 (1987).
23. Harris, M. J. & Lambert, D. L. Oxygen isotopic abundances in the atmospheres of seven red giant stars. *Astrophys. J.* **285**, 674–682 (1984).
24. Harris, M. J., Lambert, D. L. & Smith, V. V. Oxygen isotopic abundances in evolved stars. II. Eight MS and S stars. *Astrophys. J.* **299**, 375–385 (1985).
25. Dominy, J. F., Wallerstein, G. & Suntzeff, N. B. Abundances of carbon, nitrogen, and oxygen and their isotopes in the atmospheres of four SC stars. *Astrophys. J.* **300**, 325–338 (1986).
26. Harris, M. J., Lambert, D. L., Hinkle, K. H., Gustafsson, B. & Eriksson, K. Oxygen isotopic abundances in evolved stars. III. 26 Carbon stars. *Astrophys. J.* **316**, 294–304 (1987).
27. Harris, M. J., Lambert, D. L. & Smith, V. V. Oxygen isotopic abundances in evolved stars. IV. Five K giants. *Astrophys. J.* **325**, 768–775 (1988).
28. Smith, V. V. & Lambert, D. L. The chemical composition of red giants. III. Further CNO isotopic and s-process abundances in thermally pulsing asymptotic giant branch stars. *Astrophys. J. Suppl.* **72**, 387–416 (1990).
29. Kahane, C., Cernicharo, J., Gomez-Gonzales, J. & Guélin, M. Isotopic abundances in carbon-rich circumstellar envelopes: a further iteration on the oxygen isotope puzzle. *Astron. Astrophys.* **256**, 235–250 (1992).
30. Prantzos, N., Aubert, O. & Audouze, J. Evolution of the carbon and oxygen isotopes in the Galaxy. *Astron. Astrophys.* **309**, 760–774 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements A. Toppani, J. Kiener, A. Coc, N. Prantzos, M. Gounelle, V. Tatischeff, M. Chaussidon and B. Charoy are thanked for discussions, A. Kohler, B. Assouar, C. Clément and L. Marchal helped with SEM and FESEM, and T. Lhomme helped with the Raman laser. This work was supported by the Région Lorraine, and by PNP-INSU and PCMI-INSU grants.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.A. (aleon2@lnl.gov).

Mutual phase-locking of microwave spin torque nano-oscillators

Shehzaad Kaka¹, Matthew R. Pufall¹, William H. Rippard¹, Thomas J. Silva¹, Stephen E. Russek¹ & Jordan A. Katine²

The spin torque^{1,2} effect that occurs in nanometre-scale magnetic multilayer devices can be used to generate steady-state microwave signals in response to a d.c. electrical current^{3–8}. This establishes a new functionality for magneto-electronic structures that are more commonly used as magnetic field sensors and magnetic memory elements⁹. The microwave power emitted from a single spin torque nano-oscillator (STNO) is at present typically less than 1 nW. To achieve a more useful power level (on the order of microwatts), a device could consist of an array of phase coherent STNOs, in a manner analogous to arrays of Josephson junctions and larger semiconductor oscillators^{10–12}. Here we show that two STNOs in close proximity mutually phase-lock—that is, they synchronize, which is a general tendency of interacting nonlinear oscillator systems^{13–15}. The phase-locked state is distinct, characterized by a sudden narrowing of signal linewidth and an increase in power due to the coherence of the individual oscillators. Arrays of phase-locked STNOs could be used as nanometre-scale reference oscillators. Furthermore, phase control of array elements (phased array) could lead to nanometre-scale directional transmitters and receivers for wireless communications.

Mutually phase-locked interacting oscillators are surprisingly common natural occurrences. Examples of self-synchronizing systems include oscillations of interacting Josephson junctions^{10,11,16,17}, the rhythmic flashing of certain fireflies¹⁸ and the oscillations of a system of two pendulum clocks coupled through a wall, as first reported by Huygens in the seventeenth century¹⁹. Participating elements of a phase-locked system must exhibit a nonlinear response to forcing stimuli; hence, under certain conditions, a collectively ordered state emerges from a complex dynamical system. Phase-locking occurs in STNOs because magnetic precession, the source of microwave oscillations, is inherently nonlinear²⁰.

Electrical nano-contacts to thin-film magnetic bilayer mesas are d.c. current-controlled STNOs that produce microwave precession ranging from 1 GHz to beyond 40 GHz, with spectral linewidths typically in the range 2–50 MHz at room temperature^{6,21}. The oscillations are detected by measuring the time-varying voltage across the device caused by the giant magnetoresistance (GMR) effect²² and the d.c. current through the contact. When active, the STNOs are predicted to generate spinwaves flowing outward from the region immediately beneath the nano-contact²³. With the intent of using spinwave interactions to facilitate phase-locking between two STNOs, we investigated a device with two independently connected approximately 40-nm diameter contacts A and B separated by 500 nm on the same mesa (Fig. 1a, b). The two contacts are separately current biased, making each contact an independently controlled STNO. Bias-tees separate the d.c. current applied through each oscillator from the generated high-frequency output signal. The

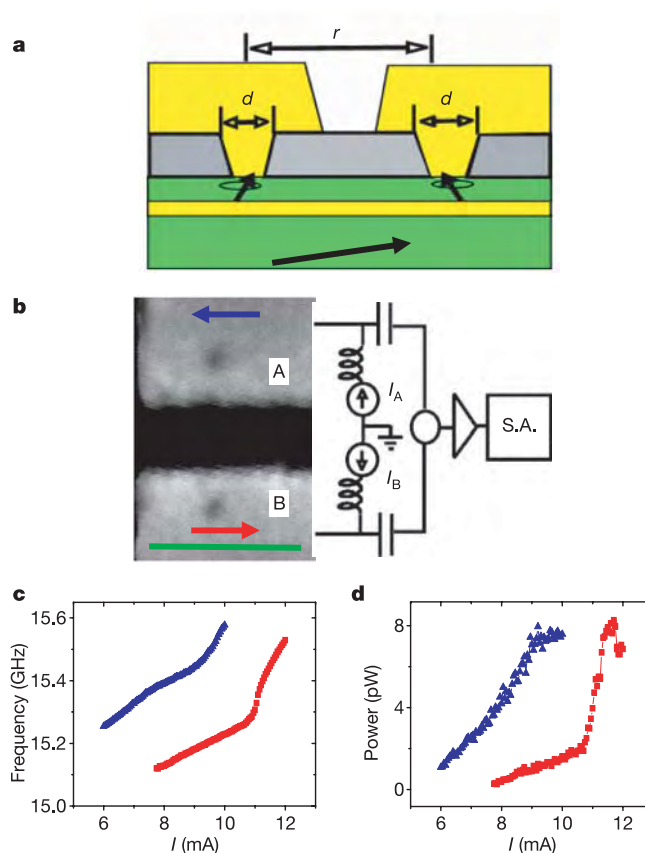


Figure 1 | Structure and basic behaviour of a two-nano-contact device.

a, Cross-sectional diagram of a two-nano-contact device structure with contact diameter $d \approx 40$ nm and contact separation $r = 500$ nm on a single mesa. The mesa layer structure is [Ta 5 nm/Cu 50 nm/Co₉₀Fe₁₀ 20 nm/Cu 5 nm/Ni₈₀Fe₂₀ 5 nm/Cu 1.5 nm/Au 2.5 nm]. **b**, Micrograph of actual two-nano-contact device with two independent leads. Scale bar (green), 500 nm. The blue arrow gives the direction of the Ampere field generated by positive current (coming out of the plane) through contact B at contact A. The red arrow is the direction of the Ampere field generated by contact A at contact B and also the direction of the in-plane component of the external magnetic field. At the right is shown a measurement diagram showing a bias tee and d.c. current source for each contact. The high-frequency power output is combined in a microwave power combiner and then sent to the spectrum analyser (S.A.). **c**, Plot of frequency of non-interacting oscillator against current. The blue curve is output for contact A, the red curve for contact B. **d**, Plot of power output against current for each non-interacting oscillator; blue triangles are for contact A, red squares for contact B.

¹Electromagnetic Technology Division, National Institute of Standards and Technology, Boulder, Colorado 80305, USA. ²Hitachi San Jose Research Center, San Jose, California 95120, USA.

output signals of both STNOs are sent to a microwave power combiner, and the combined signal is amplified and measured by a spectrum analyser (Fig. 1b). The amplifier gain has been divided out of all presented data. The electrical isolation between the contacts is -37 dB. Measurements are taken with the device placed in an external 740-mT magnetic field oriented 75° from the film plane.

The peak frequencies of each STNO, when biased alone (no current through the other contact), are shown in Fig. 1c. The frequencies exhibited by each oscillator agree with the behaviour of previously studied single-contact devices at the same applied field and field angle²¹. However, slight differences exist in the frequency and power output (Fig. 1d) for each oscillator. The frequency and power are determined from lorentzian fits to the peaks in the

measured power spectral density (spectrum). According to the data in Fig. 1c, certain combinations of currents applied to both STNOs will result in coincidence of their respective oscillation frequencies.

When the frequency of one STNO is made to approach the other, interactions cause the oscillators to lock together. Figure 2a plots the evolution of the combined spectrum from both STNOs as current I_B through contact B increases. The current I_A through contact A is fixed at 8.0 mA. As I_B increases from 7 to 8.2 mA, only signal A (sourced by contact A) is visible. The frequency f_A of A decreases slightly with I_B owing to the Ampere field (about 3 mT to 5 mT) generated by I_B . This Ampere field opposes the in-plane component of the applied field and its direction is shown by the blue arrow in Fig. 1b. For $8.2 \text{ mA} < I_B < 9.2 \text{ mA}$, the signal from contact B appears and its frequency f_B increases towards f_A with the same slope as in the non-interacting case (Fig. 1c). The spectrum at $I_B = 8.65 \text{ mA}$, shown in Fig. 2b, contains peaks from both STNOs. Above $I_B = 9.2 \text{ mA}$, f_A suddenly unites with f_B until I_B exceeds about 11 mA. These data show that both STNOs frequency-lock over a 1.5-mA range in I_B , with the implication that signals A and B are also phase-locked. We give direct evidence below of phase-locking. The spectrum of the locked state at $I_B = 9.5 \text{ mA}$ (Fig. 2c) shows a single peak with much larger amplitude and a narrower linewidth than the peaks in Fig. 2b. For $I_B \geq 11 \text{ mA}$, f_A and f_B separate and diverge. Figure 2d shows the spectrum of two peaks at $I_B = 11.5 \text{ mA}$, where both peaks are weaker and broader than the locked state peak. Unlocking occurs at a sharp jump in f_B that is also seen in the non-interacting behaviour of B at 11 mA (Fig. 1c). Figure 2e plots the linewidths as a full-width at half-maximum (FWHM) for all peaks in the spectra. During locking, the FWHM decreases by about an order of magnitude to about 2 MHz.

Next we studied the individual output of each STNO as they evolved through the locking process (in contrast to their combined signal). Figure 3a shows the evolution of only signal A as I_B is tuned and I_A is fixed at 8.0 mA. In this measurement, the high-frequency output from B is disconnected from the power combiner and terminated into a $50\text{-}\Omega$ load. The linewidth of A, shown by the

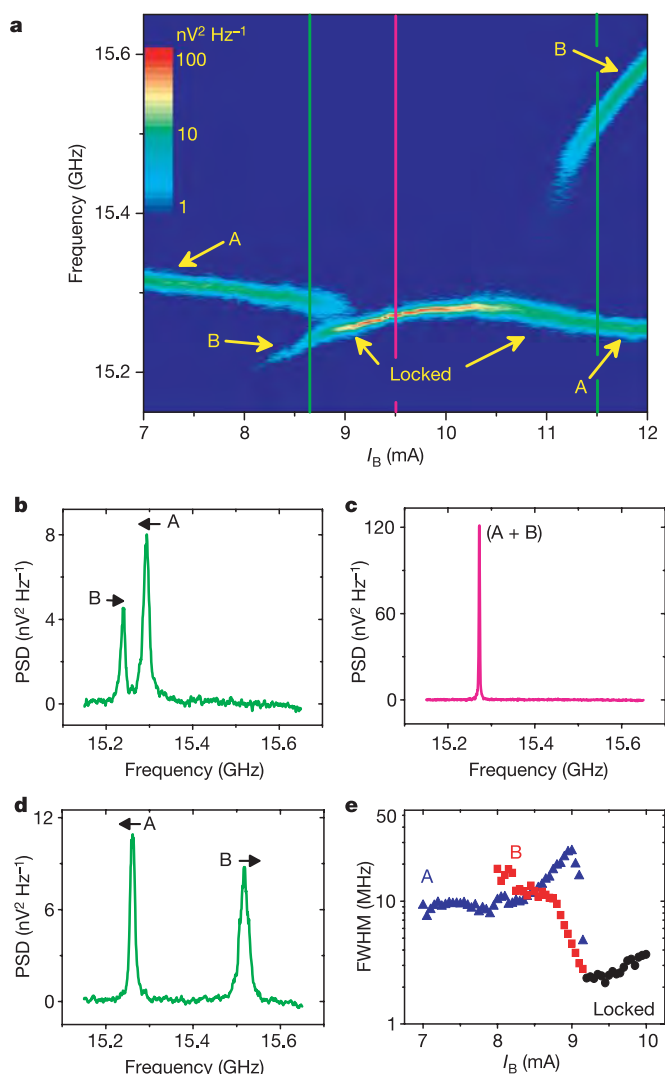


Figure 2 | Locking behaviour. **a**, Combined spectrum from both contacts as current through contact B is ramped from 7 mA to 12 mA. Current through contact A is fixed at 8 mA. Spectral intensity (colour) is a logarithmic scale. **b**, Spectrum (power spectral density; PSD) corresponding to the green vertical line in **a** at 8.65 mA. The arrows indicate the movement of the peaks as current through B increases. **c**, Spectrum corresponding to the magenta vertical line in **a** at 9.5 mA. **d**, Spectrum corresponding to the green vertical line in **a** at 11.5 mA. Arrows indicate motion of the peaks as current through contact B increases. **e**, Linewidths of combined output spectrum, where red squares correspond to the lower-frequency peak initially due to the signal from B, blue triangles correspond to the higher-frequency signal initially from A, and black circles correspond to the locked state. Uncertainty in the linewidth measurement is typically less than 0.75 MHz, which derives from one standard deviation to a lorentzian fit to the spectral peaks.

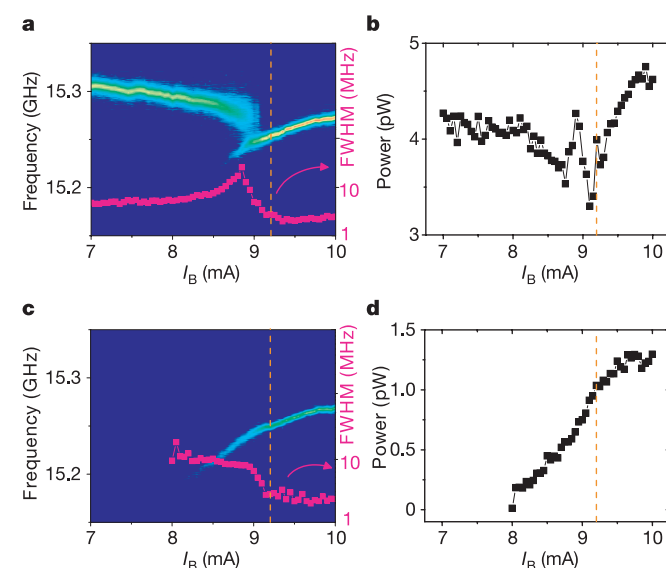


Figure 3 | Behaviour of individual oscillators. Dashed vertical lines (orange) denote the beginning of the locking range. **a**, Spectral intensity measured only for oscillator A as current through contact B is ramped and current through contact A is fixed at 8 mA. The colour scale is the same as for Fig. 2a. The superimposed magenta curve shows the linewidth of signal A on the same current scale. **b**, Power of oscillator A only. **c**, Spectral intensity measured only for oscillator B with the same currents through both contacts. The superimposed magenta curve shows the linewidth of signal B. **d**, Power of oscillator B only.

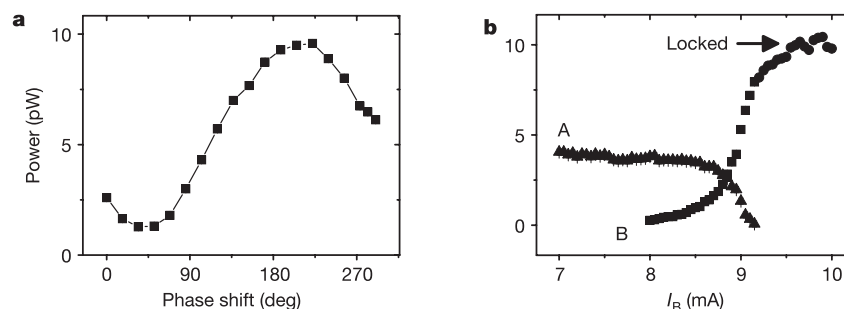


Figure 4 | Device power outputs. **a**, Power in the locked state as the phase of A is shifted. The current through contact A is 8 mA; the current through contact B is 9.55 mA. **b**, Measured power output for both oscillators as the current through contact B is ramped. The phase shifter is set for maximum output power (constructive interference) at 9.55 mA through contact

magenta curve in Fig. 3a, narrows when the system locks. The power output of A (Fig. 3b) is constant until the system nears lock; on locking, the power from A increases. The evolution of only signal B is shown in Fig. 3c, and its power output is given in Fig. 3d. The current-dependent power from B is almost identical to the non-interacting case (Fig. 1d). The dissimilarity in the power of A and B remains even as the oscillators lock. The increase in power from A during locking is possibly caused by a change to a magnetization trajectory with higher GMR as oscillator A tracks to the locked frequencies. Although the power outputs of each STNO differ, both oscillators show narrowing to about the same linewidth of 2 MHz during locking. The abrupt decrease in linewidth indicates a reduced sensitivity to input noise (such as thermal fluctuations) for both oscillators because they provide mutual feedback when locked. Such an increased stability against frequency and phase fluctuations is also known for the synchronized generators of the North American power grid¹³ and is predicted for arrays of Josephson junction oscillators¹⁶.

The phase coherence of signals A and B during locking ($I_A = 8.0$ mA and $I_B = 9.55$ mA) is shown by varying the relative phase between the outputs of the STNOs. A phase-shifting element is placed in line with the cables carrying signal A to the combiner. By fully varying the phase shifter, the phase between the two signals at the combiner is adjusted over a range of about 300° , and this produces a sinusoidal variation in the combined output power corresponding to a change from destructive to constructive interference between signal A and B (Fig. 4a). Hence, a time-independent phase relationship between the signals must exist (phase-locking occurs). With the phase shifter set to maximize the amplitude at $I_B = 9.55$ mA and $I_A = 8.0$ mA, the combined output power from both oscillators is measured as I_B is ramped (Fig. 4b). The individual oscillator powers at $I_B = 9.55$ mA are $P_A = (4.37 \pm 0.03)$ pW and $P_B = (1.19 \pm 0.01)$ pW. The measured oscillator power is proportional to the square of the oscillating voltage waveform emitted by the STNO. If each oscillator were simply locked in frequency but incoherent in phase, the time-averaged total power output would simply be $P_A + P_B = (5.56 \pm 0.04)$ pW. However, the measured output power is $P_{\text{total}} = (9.90 \pm 0.07)$ pW (Fig. 4b), which is close to the expected power for two phase-coherent voltage waveforms that constructively interfere, $P_c = P_A + P_B + 2\sqrt{(P_A P_B)} = (10.15 \pm 0.05)$ pW.

STNOs have also been shown to phase-lock (injection lock) to fixed-frequency external sources, coupled by means of input microwave currents or external microwave magnetic fields²⁴. In the system studied here, the interactions causing locking are mutual: either spinwave excitations²³ emitted by both oscillators or a.c. dipole magnetic-field interactions. Micromagnetic simulations²⁵ that incorporate a spin torque term derived by Slonczewski¹ show that large-amplitude spinwaves exist at a distance of 500 nm from the emitting nano-contact. Oscillating dipole fields, generated by localized pre-

cession underneath one contact, are estimated to be about 0.1 mT at the location of the other contact. Simulations predict this to be a sufficient field to induce locking. Moreover, a.c. magnetostatic fields caused by travelling spinwaves may exceed 10 mT at each contact, according to the micromagnetic simulations. The strength of both the spinwave and dipolar interactions, in principle, decays with the separation distance. Experimentally, for two STNOs separated by 1,000 nm, we find that although the frequencies of each oscillator can be made to intersect, weakened interactions result in no phase locking over a finite current for an applied field of 740 mT oriented 75° out of the film plane.

We have demonstrated phase-locking of two STNOs separated by 500 nm. Combined power in the phase-locked state is the coherent sum of the individual power of each oscillator. Hence, we anticipate that a phase-locked array of N STNOs can produce power that scales as N^2 , leading to substantial narrowband power output (on the order of microwatts) generated from a micrometre-sized device at room temperature. A practical STNO-array device will probably require current from a single source distributed to each oscillator through parallel or series connections. These devices have applications as sources in nanometre-scale phased arrays, which could be used in wireless chip-to-chip or intra-chip communications.

Received 12 May; accepted 14 July 2005.

- Slonczewski, J. C. Current-driven excitation of magnetic multilayers. *J. Magn. Magn. Mater.* **159**, L1–L7 (1996).
- Berger, L. Emission of spin waves by a magnetic multilayer traversed by a current. *Phys. Rev. B* **54**, 9353–9358 (1996).
- Katine, J. A., Albert, F. J., Buhrman, R. A., Myers, E. B. & Ralph, D. C. Current-driven magnetization reversal and spin wave excitations in Co/Cu/Co pillars. *Phys. Rev. Lett.* **84**, 4212–4215 (2000).
- Tsoi, M. *et al.* Generation and detection of phase-coherent current-driven magnons in magnetic multilayers. *Nature* **406**, 46–48 (2000).
- Kiselev, S. I. *et al.* Microwave oscillations of a nanomagnet driven by a spin-polarized current. *Nature* **425**, 380–383 (2003).
- Rippard, W. H., Pufall, M. R., Kaka, S., Russek, S. E. & Silva, T. J. Direct-current induced dynamics in $\text{Co}_{90}\text{Fe}_{10}/\text{Ni}_{80}\text{Fe}_{20}$ point contacts. *Phys. Rev. Lett.* **92**, 27201 (2004).
- Covington, M., Al Haj Darwish, M., Ding, Y., Gokemeijer, N. J. & Seigler, M. Current-induced magnetization dynamics in current perpendicular to the plane spin valves. *Phys. Rev. B* **69**, 184406 (2004).
- Krivorotov, I. N. *et al.* Time domain measurements of nanomagnet dynamics driven by spin-transfer torques. *Science* **307**, 228–231 (2005).
- Wolf, S. A. *et al.* Spintronics: a spin-based electronics vision for the future. *Science* **294**, 1488–1495 (2001).
- Benz, S. P. & Burroughs, C. J. Coherent emission from two dimensional Josephson junction arrays. *Appl. Phys. Lett.* **58**, 2162–2164 (1991).
- Wengler, M. J., Guan, B. & Track, E. K. 190-GHz radiation from a quasioptical Josephson junction array. *IEEE Trans. Microwave Theory Tech.* **43**, 984–988 (1995).
- Popovic, Z. B., Weikle, R. M., Kim, M. & Rutledge, D. B. A 100-MESFET planar grid oscillator. *IEEE Microwave Theory Tech.* **39**, 193–200 (1991).
- Strogatz, S. *Sync: The Emerging Science of Spontaneous Order* 51, 116 (Hyperion, New York, 2003).

14. York, R. A. Nonlinear analysis of phase relationships in quasi-optical oscillator arrays. *IEEE Trans. Microwave Theory Tech.* **41**, 1799–1809 (1993).
15. Rezavi, B. A study of injection locking and pulling in oscillators. *IEEE J. Solid State Circuits* **39**, 1415–1424 (2004).
16. Wiesenfeld, K., Colet, P. & Strogatz, S. H. Synchronization transitions in a disordered Josephson series array. *Phys. Rev. Lett.* **76**, 404–407 (1996).
17. Finnegan, T. F. & Wahlsten, S. Observation of coherent microwave radiation emitted by coupled Josephson junctions. *Appl. Phys. Lett.* **21**, 541–544 (1972).
18. Buck, J. & Buck, E. Mechanism of rhythmic synchronous flashing of fireflies. *Science* **159**, 1319–1327 (1968).
19. Bennet, M., Schatz, M. F., Rockwood, H. & Wiesenfeld, K. Huygens's clocks. *Proc. R. Soc. Lond. A* **458**, 563–579 (2002).
20. Suhl, H. The nonlinear behaviour of ferrites at high microwave signal levels. *Proc. Inst. Radio Engrs.* **44**, 1270–1284 (1956).
21. Rippard, W. H., Pufall, M. R., Kaka, S., Silva, T. J. & Russek, S. E. Current-driven microwave dynamics in magnetic point contacts as a function of applied field angle. *Phys. Rev. B* **70**, 100406 (2004).
22. Baibich, M. N. *et al.* Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices. *Phys. Rev. Lett.* **61**, 2472–2475 (1988).
23. Slonczewski, J. C. Excitation of spin waves by an electric current. *J. Magn. Magn. Mater.* **195**, L261–L268 (1999).
24. Rippard, W. H., Pufall, M. R., Kaka, S., Silva, T. J. & Russek, S. E. Injection locking and phase control of spin transfer oscillators. *Phys. Rev. Lett.* **95**, 067203 (2005).
25. Russek, S. E., McMichael, R. D., Donahue, M. J. & Kaka, S. in *Spin Dynamics in Confined Magnetic Structures II* (eds Hillebrands, B. & Ounadjela, K.) 93–156 (Springer, Berlin, 2003).

Acknowledgements We thank P. Kabos and A. Kos for assistance with microwave apparatus, and A. Slavin, M. Stiles, T. Gerrits and R. Goldfarb for discussions. This work was partly supported by the US government.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.K. (shehzu21@gmail.com).

Phase-locking in double-point-contact spin-transfer devices

F. B. Mancoff¹, N. D. Rizzo¹, B. N. Engel¹ & S. Tehrani¹

Spin-transfer^{1,2} in nanometre-scale magnetic devices results from the torque on a ferromagnet owing to its interaction with a spin-polarized current and the electrons' spin angular momentum. Experiments have detected either a reversal^{3–16} or high-frequency (GHz) steady-state precession^{17–23} of the magnetization in giant magnetoresistance spin valves and magnetic tunnel junctions with current densities of more than 10^7 A cm^{-2} . Spin-transfer devices may enable high-density, low-power magnetic random access memory^{24,25} or direct-current-driven nanometre-sized microwave oscillators. Here we show that the magnetization oscillations induced by spin-transfer in two 80-nm-diameter giant-magnetoresistance point contacts in close proximity to each other can phase-lock into a single resonance over a frequency range from approximately <10 to >24 GHz for contact spacings of less than about ~200 nm. The output power from these contact pairs with small spacing is approximately twice the total power from more widely spaced (~400 nm and greater) contact pairs that undergo separate resonances, indicating that the closely spaced pairs are phase-locked with zero phase shift. Phase-locking may enable control of large arrays of coupled spin-transfer devices with increased power output for microwave oscillator applications.

We measured the room-temperature, high-frequency spin-transfer-induced oscillations in devices with two nearby point contacts to a giant magnetoresistance (GMR) thin-film spin valve (schematic cross-section in Fig. 1a). The sputter-deposited film consisted of a 5-nm-Pd/25-nm-Cu base electrode, a 20-nm- $\text{Co}_{81}\text{Fe}_{19}$ fixed magnetic layer, a 6-nm-Cu spacer, a 4.5-nm- $\text{Ni}_{80}\text{Fe}_{20}$ free magnetic layer, and a 2-nm-Cu/3.5-nm-Pd cap. The GMR film was patterned to sizes as small as only $8 \mu\text{m}$, and the point contacts were fabricated by electron beam lithography followed by reactive ion etching through an SiO_2 insulating layer to the GMR film. A single top electrode connected electrically to the two point contacts in parallel. Figure 1b shows a cross-section scanning electron microscope (SEM) image of the contacts, each 80 nm in diameter, with 120 nm centre-to-centre spacing, while Fig. 1c is a top-down SEM image of a similar sample with 240-nm spacing.

For high-frequency measurements, a total direct current $I_{\text{d.c.}}$ was applied to the contact pair through a bias tee. Spin-transfer oscillations excited by $I_{\text{d.c.}}$ created voltage oscillations through the GMR that were measured with a spectrum analyser through the other port of the bias tee. Positive $I_{\text{d.c.}}$ was defined as electron flow from the thin, $\text{Ni}_{80}\text{Fe}_{20}$ free layer to the thick, $\text{Co}_{81}\text{Fe}_{19}$ fixed layer. Spin-transfer torque excited precession of the $\text{Ni}_{80}\text{Fe}_{20}$ magnetization, while the $\text{Co}_{81}\text{Fe}_{19}$ magnetization remained nearly fixed owing to its greater thickness and magnetization. Measurements on similar devices with only a single GMR point contact (F.B.M., N.D.R., B.N.E. & S.T., manuscript in preparation) showed output power up to ~200 pW, precession linewidths down to <10 MHz, and ratios of peak frequency to linewidth up to ~2,000.

Figure 2 shows the measured frequency spectrum voltage amplitude versus $I_{\text{d.c.}}$ for three devices with centre-to-centre spacings of 800 nm (Fig. 2a), 150 nm (Fig. 2b), and 120 nm (Fig. 2c). All the data in this paper was taken in a magnetic field of between 0.8–1.2 T applied perpendicular to the plane of the film. The spin-transfer precession, or resonance, appeared as a frequency spectrum peak at a given $I_{\text{d.c.}}$. The peak frequencies for all spacings increased approximately linearly with increased $I_{\text{d.c.}}$, over a frequency range of <10 to >24 GHz, in agreement with similar measurements on single GMR point contacts (F.B.M., N.D.R., B.N.E. & S.T., manuscript in preparation; ref. 20). The number and behaviour of the observed spin-transfer peaks depended on the two contacts' relative spacing. For the widely spaced device (Fig. 2a) with 800-nm spacing, we observed two separate peaks for nearly all $I_{\text{d.c.}}$. The two peaks had comparable resonance frequencies that increased at similar average rates versus $I_{\text{d.c.}}$. Each peak displayed frequency jumps at multiple $I_{\text{d.c.}}$ values, similar to discontinuities observed for single point contacts (F.B.M., N.D.R., B.N.E. & S.T., manuscript in preparation; ref. 20). The two sets of jumps were uncorrelated, so the pair of contacts at 800-nm spacing was undergoing independent, uncoupled spin-transfer oscillations.

In contrast, the closely spaced contact pair (Fig. 2c) at 120-nm spacing displayed only a single peak in frequency for all $I_{\text{d.c.}}$. The change from two independent peaks at 800-nm spacing to a single peak at 120-nm spacing indicates that interactions between the point contacts at closer spacing coupled the spin-transfer oscillations. The coupling strength is sufficient to bring together in frequency the resonances from the two different contacts that would otherwise be separated by several hundred megahertz for uncoupled devices, as

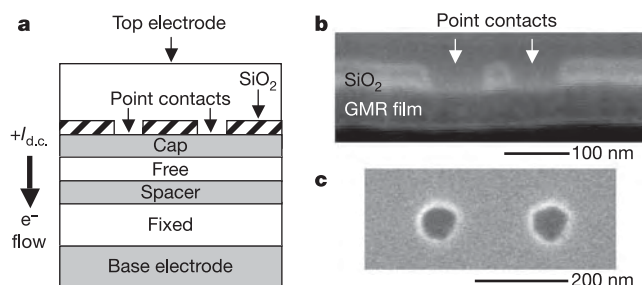


Figure 1 | Overview of double-point-contact device. **a**, Schematic cross-section of two point contacts to a GMR spin valve film through an SiO_2 insulator. A single top electrode was connected electrically to the two contacts. The spin valve contained free and fixed magnetic layers with a nonmagnetic spacer, base electrode and cap. Positive direct current flow $+I_{\text{d.c.}}$ was defined as electrons flowing from the free to the fixed layer. **b**, Cross-section SEM image of two 80-nm-diameter contacts with a 120 nm centre-to-centre spacing. **c**, Top-down SEM image of two 80-nm contacts with a 240-nm spacing.

¹Technology Solutions Organization, Freescale Semiconductor Inc., Chandler, Arizona 85224, USA.

in Fig. 2a. For the device with somewhat greater spacing (150 nm, Fig. 2b), we detected two separate peaks for most $I_{d.c.}$. However, here the two peaks progressed nearly parallel in frequency versus $I_{d.c.}$, compared to the two independent sets of frequency jumps for the 800-nm spacing device (Fig. 2a). Thus, the 150-nm spacing data in Fig. 2b behaved with a coupling strength intermediate between the two independent peaks at 800-nm spacing and the single peak at 120-nm spacing.

We observed a significant difference in power output between devices with a relatively close spacing and a single resonance peak compared to devices with larger spacing and two independent peaks. Figure 2d shows that the typical integrated power normalized by $I_{d.c.}^2$ for the single peak in a device with 120-nm spacing (black crosses) was clearly greater than the combined power for the two peaks in a device with 800-nm spacing (red squares). Of seven devices (with spacings of 120–200 nm) showing a single peak similar to Fig. 2c, the average output power was $11 \pm 2 \text{ nW A}^{-2}$ (where specified uncertainty is the standard error) in a 1-T applied magnetic field. In contrast, for all seven devices (with spacings of 400 and 800 nm) showing two independent peaks similar to Fig. 2a, the average output power was $5 \pm 0.5 \text{ nW A}^{-2}$ for the two peaks combined in a 1-T magnetic field. A similar factor-of- ~ 2 difference in power was also observed in 1.2 T, with $8 \pm 1 \text{ nW A}^{-2}$ for devices with small spacing and one peak and $4 \pm 0.5 \text{ nW A}^{-2}$ for devices with large spacing and two peaks.

The factor-of- ~ 2 difference in output power indicates that the closely spaced double-point contacts were phase-locked. We modelled the two contacts as resistors in parallel with each other and in series with a single current source $I_{d.c.}$. The two resistances oscillate in time with a phase difference ϕ and equal amplitude and

frequency. For two uncoupled contacts oscillating independently with random ϕ , as expected in Fig. 2a, the modelled total power is half the power of the case where $\phi = 0$ for two in-phase oscillators. Thus, we conclude that the two point-contact oscillators with close intercontact spacing (Fig. 2c) are phase-locked with $\phi = 0$, while those with wide spacing (Fig. 2a) are uncoupled.

We measured spin transfer in over 30 devices with various spacings and repeatedly observed the trend in Figs 2a–c. The statistical distributions for these measurements are shown in Fig. 3, where we plot histograms for the three types of spin-transfer frequency-domain behaviour (two peaks, intermediate, and one peak) versus intercontact spacing. The solid lines are guides to the eye. The ‘two peaks’ histogram (solid diamonds) is maximum for the largest spacings of 400 nm and more, and drops to zero at small spacings. These data correspond to devices with two spin-transfer peaks observed at each $I_{d.c.}$ and with each peak showing its own independent frequency jumps, indicating that the point-contact pair was uncoupled. The ‘one peak’ histogram (open squares) is maximum at the smallest spacings of 120–150 nm and decreases to zero above 200 nm. These data correspond to devices with a single resonance peak observed at each $I_{d.c.}$, indicating strong coupling and associated phase-locking between the oscillators. Finally, the ‘intermediate’ histogram (solid triangles) shows a third category of spin-transfer data with a maximum around a spacing of 200 nm. These data correspond to devices which showed a transition from one peak to two within an $I_{d.c.}$ sweep, or a single peak over most $I_{d.c.}$ values with an occasional weaker second peak, or two peaks moving nearly parallel in frequency with $I_{d.c.}$, as in Fig. 2b. This intermediate category represents devices in a transition range between those with strong coupling with phase-locking and those with very weak or no coupling between the two independent oscillators.

Figure 3 indicates strong phase-locking in our double-point contacts at centre-to-centre spacings of less than ~ 200 nm. The effective area of each point contact undergoing spin-transfer oscillations exceeds the physical contact area owing to coupling to the extended $\text{Ni}_{80}\text{Fe}_{20}$ free layer magnetic film surrounding the contact. For a single contact of nominally 80 nm in diameter, we determined the effective region of magnetic excitation to be approximately 170 nm in diameter (F.B.M., N.D.R., B.N.E. & S.T., manuscript in preparation). Thus, the magnetic excitation areas of a pair of

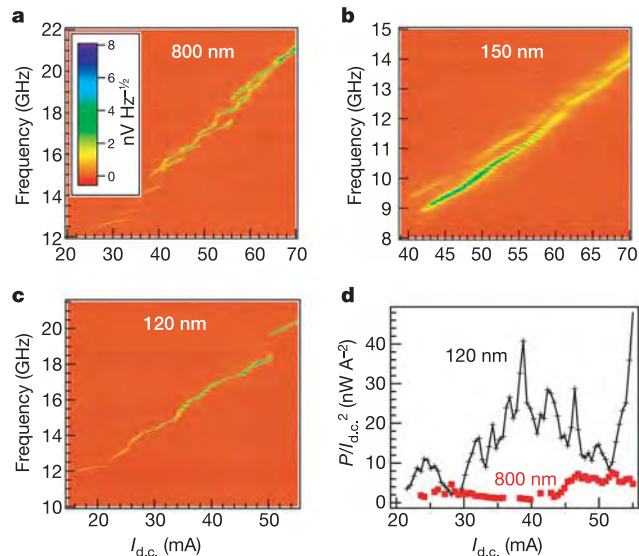


Figure 2 | High-frequency output for devices with two 80-nm point contacts and varied intercontact spacing. Identical colour scales are used in **a**, **b** and **c**. **a**, Map of spectrum amplitude versus frequency and applied current bias $I_{d.c.}$ for a double-point contact with 800-nm centre-to-centre spacing. Two spin-transfer peaks, each with its own set of frequency jumps, were detected. **b**, Map of spectrum amplitude versus frequency and $I_{d.c.}$ for a double-point contact with 150-nm spacing. Two spin-transfer peaks moving approximately parallel in frequency with $I_{d.c.}$ were observed for nearly the whole $I_{d.c.}$ range. **c**, Map of spectrum amplitude versus frequency and $I_{d.c.}$ for a double-point contact with 120-nm spacing. A single spin-transfer resonance peak was observed for all $I_{d.c.}$, and the frequency increased roughly linearly with $I_{d.c.}$. **d**, Integrated power P normalized by $I_{d.c.}^2$ versus $I_{d.c.}$ for double-point-contact devices with 120-nm (black crosses) and 800-nm (red squares) spacings. Corresponding frequency spectra show either one or two spin-transfer peaks respectively. The curve for the 800 nm spaced sample includes the sum of the power in its two peaks.

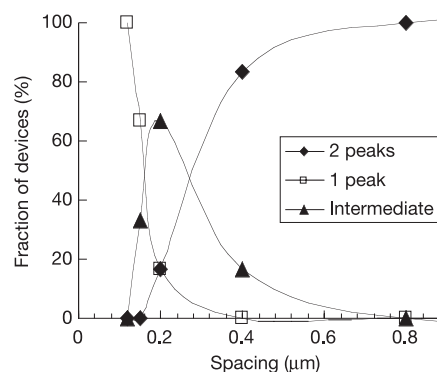


Figure 3 | Histograms showing statistical distribution of spin-transfer behaviour in devices with two 80-nm-diameter contacts and varied centre-to-centre contact spacing. Square data points, showing the number of devices for which a single spin-transfer peak was observed at all values of $I_{d.c.}$ as in Fig. 2c, decrease sharply from a maximum at the smallest spacing of 120 nm. Diamond data points, showing the number of devices for which two spin-transfer peaks were observed, each with its own frequency jumps versus $I_{d.c.}$ as in Fig. 2a, increase to a maximum at large spacings > 400 nm. Triangle data points show the number of devices with intermediate spin-transfer characteristics, such as one large spin-transfer peak and a second smaller one for some $I_{d.c.}$ values, or a transition from one peak to two within an $I_{d.c.}$ sweep, or two peaks moving nearly parallel in frequency with $I_{d.c.}$ similar to Fig. 2b. Solid lines are guides to the eye.

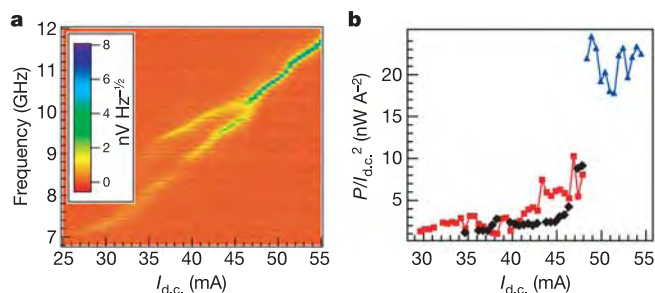


Figure 4 | High-frequency output for two 80-nm point contacts with 150-nm intercontact spacing. **a**, Map of spectrum amplitude versus frequency and applied current bias $I_{d.c.}$ plotted on an identical colour scale to Fig. 2. Two spin-transfer peaks were observed at low $I_{d.c.}$ below ~ 47 mA, and only a single peak was observed for larger $I_{d.c.}$. **b**, Integrated power P normalized by $I_{d.c.}^2$ versus $I_{d.c.}$ for the frequency spectra in **a**. The data for $I_{d.c.}$ below the transition at ~ 47 mA is given by the red squares and black diamonds for the lower and higher frequency peaks respectively over this $I_{d.c.}$ range. Blue triangles show the data for the single peak observed at larger $I_{d.c.}$. The transition from two peaks to one was accompanied by an increase in the total normalized power $P/I_{d.c.}^2$ by a factor of ~ 2 , indicating that phase-locking between the resonances had occurred.

nominally 80 nm in diameter contacts overlap for centre-to-centre spacings < 170 nm, consistent with the ~ 200 nm spacing length scale in Fig. 3 being the onset of strong phase-locking. We estimate that exchange coupling is a stronger interaction mechanism than dipole-dipole magnetic fields, although the two effects are of the same order of magnitude in energy cost.

Figure 4 gives another example of spin-transfer oscillations in a double-point-contact device with 150 nm intercontact spacing and an intermediate coupling strength. For the frequency spectra map versus $I_{d.c.}$ in Fig. 4a, we observed the onset of one spin-transfer peak for $I_{d.c.} > \sim 28$ mA followed by a second peak of comparable amplitude for $I_{d.c.} > \sim 35$ mA. The two separate peaks then co-existed for $I_{d.c.}$ up to ~ 47 mA followed by a transition to a single peak above this current, probably indicating that the contacts have phase-locked. Figure 4b shows the corresponding integrated output power normalized by $I_{d.c.}^2$. The blue triangles are for the single peak at high $I_{d.c.}$, while the red squares and black diamonds are the individual powers in the two peaks at lower $I_{d.c.}$, with the red squares (black diamonds) showing the data for the lower (higher) frequency peak of these two. We observed an average power of ~ 5 nW A $^{-2}$ in each of the two peaks with a total ~ 10 nW A $^{-2}$ just below the transition at $I_{d.c.} \approx 47$ mA, as compared to ~ 20 nW A $^{-2}$ in the single peak at higher current. This factor-of-2 power increase indicates that the two contacts have phase-locked with $\phi = 0$, as discussed above. Thus, Fig. 4 illustrates an electrically tunable transition of the double-point-contact device into and out of phase-locking. For this device, the linewidth full-width-at-half-maximum (FWHM) decreases across the transition, from ~ 200 MHz for each of the two peaks at low $I_{d.c.}$, narrowing to ~ 80 MHz for the single peak at high $I_{d.c.}$.

A decreased linewidth upon phase-locking is expected if thermal fluctuations contribute significantly. Then for the phase-locked pair, thermal energy fluctuations act on a device with greater magnetic volume, so the effective fluctuation field and FWHM will decrease²⁶, as in Fig. 4. On average, we observed a slight (~ 10 – 30%) decrease in the FWHM for the closely spaced devices with one peak compared to ~ 200 MHz FWHM for widely spaced devices with two peaks.

Thus we have showed that two nanosized spin-transfer oscillators will phase-lock to each other when the intercontact spacing is less than ~ 200 nm. We observed only one resonance for closely spaced contacts and found that the average power was approximately twice the total for two separate resonances at large spacings, indicating that phase-locking had occurred for the closely spaced contacts. Such phase-locking offers a way of controlling larger arrays of spin-transfer

oscillators and increasing the output power for tunable, microwave oscillator device applications.

Received 31 May; accepted 13 July 2005.

- Slonczewski, J. C. Current-driven excitation of magnetic multilayers. *J. Magn. Magn. Mater.* **159**, L1–L7 (1996).
- Berger, L. Emission of spin waves by a magnetic multilayer traversed by a current. *Phys. Rev. B* **54**, 9353–9358 (1996).
- Myers, E. B., Ralph, D. C., Katine, J. A., Louie, R. N. & Buhrman, R. A. Current-induced switching of domains in magnetic multilayer devices. *Science* **285**, 867–870 (1999).
- Wegrowe, J.-E., Kelly, D., Jaccard, Y., Guittienne, Ph. & Ansermet, J.-Ph. Current-induced magnetization reversal in magnetic nanowires. *Europhys. Lett.* **45**, 626–632 (1999).
- Katine, J. A., Albert, F. J., Buhrman, R. A., Myers, E. B. & Ralph, D. C. Current-driven magnetization reversal and spin-wave excitations in Co/Cu/Co pillars. *Phys. Rev. Lett.* **84**, 3149–3152 (2000).
- Grollier, J. et al. Spin-polarized current induced switching in Co/Cu/Co pillars. *Appl. Phys. Lett.* **78**, 3663–3665 (2001).
- Sun, J. Z., Monsma, D. J., Abraham, D. W., Rooks, M. J. & Koch, R. H. Batch-fabricated spin-injection magnetic switches. *Appl. Phys. Lett.* **81**, 2202–2204 (2002).
- Urazhdin, S., Birge, N. O., Pratt, W. P. Jr & Bass, J. Current-driven magnetic excitations in permalloy-based multilayer nanopillars. *Phys. Rev. Lett.* **91**, 146803 (2003).
- Mancoff, F. B. et al. Angular dependence of spin-transfer switching in a magnetic nanostructure. *Appl. Phys. Lett.* **83**, 1596–1598 (2003).
- Huai, Y., Albert, F., Nguyen, P., Pakala, M. & Valet, T. Observation of spin-transfer switching in deep submicron-sized and low-resistance magnetic tunnel junctions. *Appl. Phys. Lett.* **84**, 3118–3120 (2004).
- Ozyilmaz, B. et al. Current-induced magnetization reversal in high magnetic fields in Co/Cu/Co nanopillars. *Phys. Rev. Lett.* **91**, 067203 (2003).
- Yagami, K., Tulapurkar, A. A., Fukushima, A. & Suzuki, Y. Low-current spin-transfer switching and its thermal durability in a low-saturation-magnetization nanomagnet. *Appl. Phys. Lett.* **85**, 5634–5636 (2004).
- Jiang, Y. et al. Substantial reduction of critical current for magnetization switching in an exchange-biased spin valve. *Nature Mater.* **3**, 361–363 (2004).
- Chen, T. Y., Ji, Y., Chien, C. L. & Stiles, M. D. Current-driven switching in a single exchange-biased ferromagnetic layer. *Phys. Rev. Lett.* **93**, 026601 (2004).
- Lacour, D., Katine, J. A., Smith, N., Carey, M. J. & Childress, J. R. Thermal effects on the magnetic-field dependence of spin-transfer-induced magnetization reversal. *Appl. Phys. Lett.* **85**, 4681–4683 (2004).
- Lee, K. J. et al. Spin transfer effect in spin-valve pillars for current-perpendicular-to-plane magnetoresistive heads. *J. Appl. Phys.* **95**, 7423–7428 (2004).
- Tsoi, M. et al. Generation and detection of phase-coherent current-driven magnons in magnetic multilayers. *Nature* **406**, 46–48 (2000).
- Kiselev, S. I. et al. Microwave oscillations of a nanomagnet driven by a spin-polarized current. *Nature* **425**, 380–383 (2003).
- Covington, M., AlHajDarwish, M., Ding, Y., Gokemeijer, N. J. & Seigler, M. A. Current-induced magnetization dynamics in current perpendicular to the plane spin valves. *Phys. Rev. B* **69**, 184406 (2004).
- Rippard, W. H., Pufall, M. R., Kaka, S., Russek, S. E. & Silva, T. J. Direct-current induced dynamics in Co₉₀Fe₁₀ / Ni₈₀Fe₂₀ point contacts. *Phys. Rev. Lett.* **92**, 027201 (2004).
- Kiselev, S. I. et al. Current-induced nanomagnet dynamics for magnetic fields perpendicular to the sample plane. *Phys. Rev. Lett.* **93**, 036601 (2004).
- Rippard, W. H., Pufall, M. R., Kaka, S., Silva, T. J. & Russek, S. E. Current-driven microwave dynamics in magnetic point contacts as a function of applied field angle. *Phys. Rev. B* **70**, 100406 (2004).
- Krivorotov, I. N. et al. Time-domain measurements of nanomagnet dynamics driven by spin-transfer torques. *Science* **307**, 228–231 (2005).
- Engel, B. N. et al. A 4-Mbit toggle MRAM based on a novel bit and switching method. *IEEE Trans. Mag.* **41**, 132–136 (2005).
- Parkin, S. S. P. et al. Exchange-biased magnetic tunnel junctions and application to nonvolatile magnetic random access memory. *J. Appl. Phys.* **85**, 5828–5833 (1999).
- Russek, S. E., Kaka, S., Rippard, W. H., Pufall, M. R. & Silva, T. J. Finite-temperature modeling of nanoscale spin-transfer oscillators. *Phys. Rev. B* **71**, 104425 (2005).

Acknowledgements We thank W. H. Rippard, T. J. Silva and S. E. Russek for discussions. This work was supported in part by the DARPA SPINS programme through Motorola.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.B.M. (fred.mancoff@freescale.com).

LETTERS

Astronomical pacing of methane release in the Early Jurassic period

David B. Kemp¹, Angela L. Coe¹, Anthony S. Cohen¹ & Lorenz Schwark²

A pronounced negative carbon-isotope ($\delta^{13}\text{C}$) excursion of $\sim 5\text{--}7\text{‰}$ (refs 1–7) indicates the occurrence of a significant perturbation to the global carbon cycle during the Early Jurassic period (early Toarcian age, ~ 183 million years ago). The rapid release of ^{12}C -enriched biogenic methane as a result of continental-shelf methane hydrate dissociation has been put forward as a possible explanation for this observation^{1,7,8}. Here we report high-resolution organic carbon-isotope data from well-preserved mudrocks in Yorkshire, UK, which demonstrate that the carbon-isotope excursion occurred in three abrupt stages, each showing a shift of -2‰ to -3‰ . Spectral analysis of these carbon-isotope measurements and of high-resolution carbonate abundance data reveals a regular cyclicity. We interpret these results as providing strong evidence that methane release proceeded in three rapid pulses and that these pulses were controlled by astronomically forced changes in climate, superimposed upon longer-term global warming. We also find that the first two pulses of methane release each coincided with the extinction of a large proportion of marine species⁹.

For a brief interval during the early Toarcian, the Earth experienced a severe environmental crisis. This crisis is marked in lower Toarcian strata by a pronounced negative excursion of $\sim 5\text{--}7\text{‰}$ in the $\delta^{13}\text{C}$ of all the major biospheric reservoirs of carbon (marine organic matter^{1–7}, marine carbonate^{2–6} and terrestrial plant material⁷), globally elevated levels of organic carbon burial^{10,11}, evidence for a sudden rise in seawater palaeotemperatures^{12,13}, evidence for a 400–800% increase in global weathering rates¹, and the mass extinction of terrestrial¹⁴ and marine organisms^{9,15}. Exceptionally well-preserved, unbioturbated and laminated organic-rich mudrocks that were deposited on the continental shelf and are now exposed in Yorkshire, UK, provide the most stratigraphically complete and accurate record of this global event. In this study we have characterized the duration and precise structure of this environmental crisis at high resolution using carbon-isotope analyses of bulk organic carbon ($\delta^{13}\text{C}_{\text{org}}$) from 449 samples collected over a 12 m interval of these rocks (Fig. 1).

Our data reveal that there are three clearly defined, abrupt negative shifts in $\delta^{13}\text{C}_{\text{org}}$ in the uppermost *Dactylioceras semicelatum* ammonite subzone (*Dactylioceras tenuicostatum* zone) and the lowermost *Harpoceras exaratum* subzone (*Harpoceras falciferum* zone). The first two abrupt negative shifts are $\sim 2\text{‰}$ each (Fig. 1, shifts A and B), and are separated by a ~ 70 cm interval with a less variable signal. Each abrupt shift occurs over 7.5 cm of strata. Above B, there is an increase of $\sim 3\text{‰}$ to less negative values over ~ 110 cm before the third and final abrupt negative shift of $\sim 2\text{‰}$ that occurs over 2.5 cm (Fig. 1, shift C). Above shift C, the $\delta^{13}\text{C}_{\text{org}}$ signal shows low amplitude, high frequency fluctuations superimposed on an overall stabilization and return to near pre-excursion values over ~ 6 m of strata. Clearly

defined symmetrical fluctuations in $\delta^{13}\text{C}_{\text{org}}$ are present before the first of the abrupt shifts (that is, before shift A) in the *semicelatum* subzone, which are superimposed upon a decreasing trend (Fig. 1). These $\delta^{13}\text{C}_{\text{org}}$ fluctuations, and the stratigraphic position of the three abrupt shifts, correspond closely in both wavelength and phase relationship to cyclical variations in CaCO_3 concentration measured over the same interval (Fig. 1). Analysis of the CaCO_3 through the studied section reveals that it is primary in origin and is derived from both macro- and micro-fossil shell material.

Lower resolution $\delta^{13}\text{C}$ data of bulk organic carbon from lower Toarcian rocks of the Belluno trough, Italy⁶, show an excursion pattern that is similar to the structure defined by our high-resolution $\delta^{13}\text{C}_{\text{org}}$ data from Yorkshire. In addition, the Belluno trough section was suitable for $\delta^{13}\text{C}$ analysis of carbonate ($\delta^{13}\text{C}_{\text{carb}}$) and these data clearly show that the $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{13}\text{C}_{\text{org}}$ signals are synchronous and that they possess the same structure⁶. $\delta^{13}\text{C}$ measurements of specific organic compounds from contemporaneous lower Toarcian mudrocks in Germany³ reveal a pronounced negative excursion of similar magnitude to the bulk organic carbon in all the analysed organic compounds. These compound-specific changes in $\delta^{13}\text{C}$, coupled with the presence of a synchronous $\delta^{13}\text{C}_{\text{carb}}$ excursion^{2–6}, demonstrate unequivocally that the early Toarcian isotopic changes were not caused by changes in the balance of terrestrial and marine organic matter, or by changes in the physiology of the photoautotrophic community responsible for marine organic matter production³ (see Supplementary Information).

Spectral analysis of all our new $\delta^{13}\text{C}_{\text{org}}$ and CaCO_3 data reveals, in both cases, the presence of a regular cyclicity with a mean wavelength of 81 cm in the lower part of the section (Fig. 2). Radiometric ages from correlative sections have dated the base of the *tenuicostatum* zone at $183.6^{+1.7}_{-1.1}$ Myr and the base of the *Haugia variabilis* zone (which occurs above the *falciferum* zone) at 181.4 ± 1.2 Myr (ref. 16). On the basis of these radiometric ages and the thickness of the succession spanning these biozones in Yorkshire, each 81 cm cycle represents 21.46 kyr. This independent estimate of time strongly suggests that the $\delta^{13}\text{C}_{\text{org}}$ and CaCO_3 cycles represent astronomically forced precession cycles, which have a mean duration of 21 kyr. Taking the extreme errors of these dates, it is possible that the cycles represent 40 kyr astronomically forced obliquity cycles. However, precession, rather than obliquity, would have been the dominant astronomical climate forcing parameter at the palaeolatitude where these rocks were deposited ($\sim 30^\circ\text{N}$)¹⁷. In addition, contemporaneous strata¹¹ deposited at similar palaeolatitudes show that precession-forced CaCO_3 productivity cycles dominate sedimentation patterns over the same stratigraphic interval as that covered by our data, whereas there is no evidence for a strong obliquity component¹⁷. Furthermore, a previous estimate of the duration of this part of the Yorkshire section, which used the presence of 50 μm

¹Department of Earth Sciences, Centre for Earth, Planetary, Space & Astronomical Research, The Open University, Milton Keynes MK7 6AA, UK. ²Geologisches Institut, Universität zu Köln, Zùlpicher Strasse 49a, D-50674 Köln, Germany.

lamina couplets that were interpreted to represent annual varves^{18,19}, is in close agreement with the duration obtained by interpreting the 81 cm cycles as representing precession. Assigning a precession-controlled origin to these cycles allows us to construct a floating astronomical timescale which constrains the duration of the abrupt $\delta^{13}\text{C}_{\text{org}}$ shifts A, B and C to <2 kyr each, with all three shifts occurring within a ~ 60 kyr period (Fig. 1).

The brevity and structure of the $\delta^{13}\text{C}_{\text{org}}$ shifts A, B and C provide unequivocal evidence for isotopically light carbon entering the global exchangeable carbon reservoir in three rapid pulses. The only currently known source of ^{13}C -depleted carbon that could cause three successive shifts of such magnitude is from methane hydrates ($\delta^{13}\text{C}_{\text{methane}} \approx -60\text{‰}$), thus adding further weight to the hydrate dissociation hypothesis originally proposed⁷ for the early Toarcian event. The stratigraphic relationship between these pulses and the cyclostratigraphy established here permits a direct causal link to be made between methane hydrate dissociation and astronomical forcing. Furthermore, the stratigraphic positions of shifts A and B

coincide respectively with the abrupt loss of first 67%, and then a further 50%, of marine invertebrate species (benthic and planktonic) from the same Yorkshire succession⁹ (Fig. 1), strongly suggesting that these marine extinction events were a direct consequence of astronomically forced methane release. The temporary disappearance and selective extinction of many species of marine phytoplankton over the same interval indicates that a major crisis affected primary producers¹⁵. An appreciable proportion of terrestrial organisms also became extinct during the Toarcian¹⁴. Terrestrial extinction was likely to have been related to the immediate effects of methane hydrate dissociation and to the sudden, severe perturbation to the global carbon cycle following the introduction of methane to the atmosphere and its rapid oxidation to CO_2 . Seawater palaeotemperatures estimated from Mg/Ca ratios and $\delta^{18}\text{O}$ analyses of belemnites from the Yorkshire section indicate that, within the stratigraphic resolution of these belemnite analyses, a pronounced and sudden seawater warming of $7\text{--}13^\circ\text{C}$ (refs 12, 13) occurred at the stratigraphic position of the $\delta^{13}\text{C}_{\text{org}}$ shift A. We attribute this

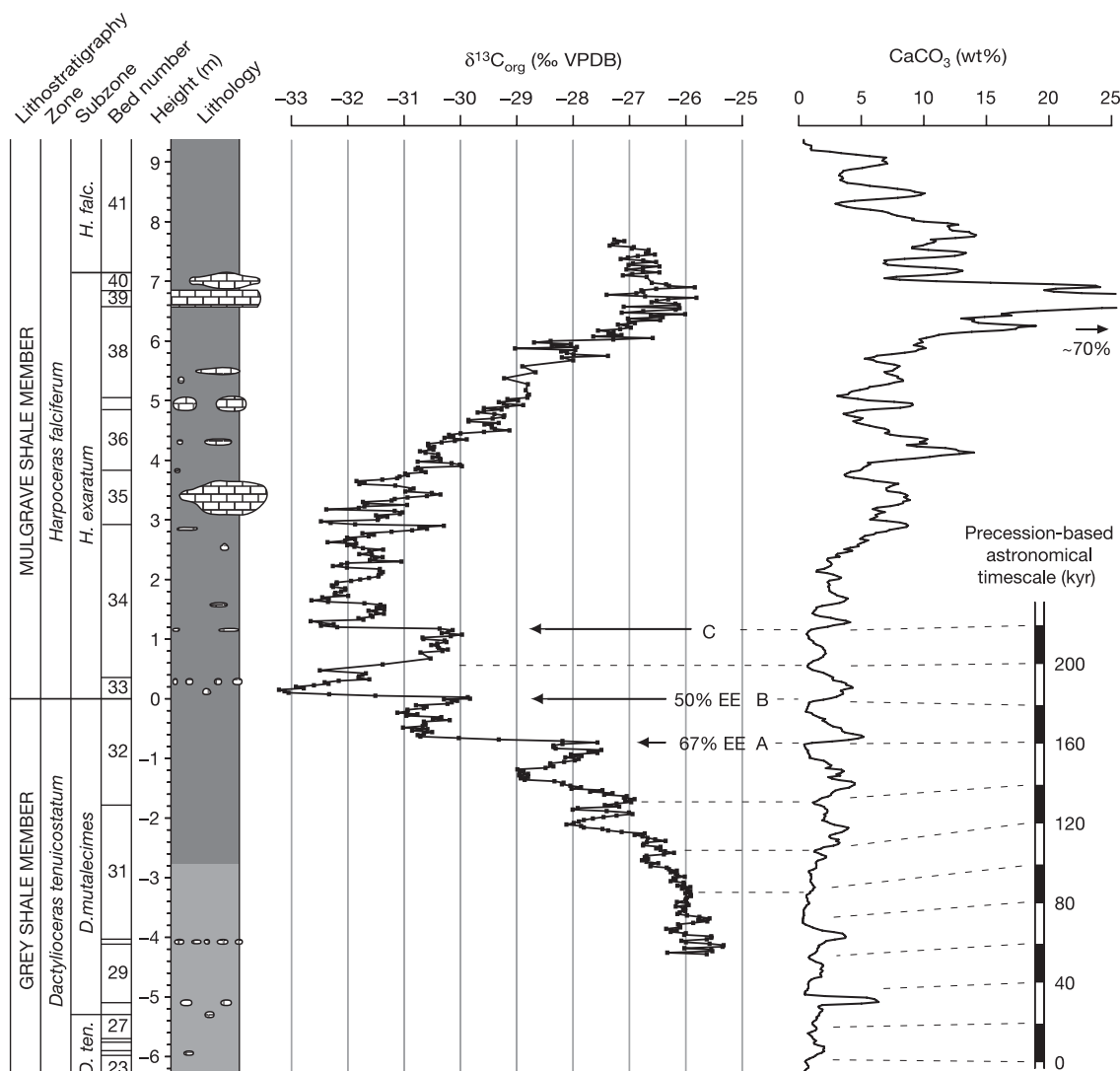


Figure 1 | $\delta^{13}\text{C}_{\text{org}}$ and CaCO_3 data through lower Toarcian sedimentary deposits. Data are from Hawsker Bottoms and Port Mulgrave in Yorkshire; CaCO_3 concentration is given as a five-point moving average. For the longer term variations in $\delta^{13}\text{C}_{\text{org}}$ ($\delta^{13}\text{C} = 1,000 \times [(^{13}\text{C}/^{12}\text{C})_{\text{sample}} / (^{13}\text{C}/^{12}\text{C})_{\text{standard}} - 1]$) from the same sections, see ref. 1 and Supplementary Information. Ammonite biostratigraphy and bed numbers are from ref. 28. Lithologies in the section are dark-grey mudrocks (shown by dark-grey shading), medium-grey mudrocks (pale-grey shading) and carbonate bands

and nodules (brick pattern). Abrupt shifts A, B and C in $\delta^{13}\text{C}_{\text{org}}$ are interpreted to represent astronomically forced pulses of methane hydrate dissociation (note the constant phase relationship between the $\delta^{13}\text{C}_{\text{org}}$ shifts and the carbonate cycles and the lack of an abrupt shift between B and C). A and B coincide with two marine extinction events⁹ (EE) involving the loss of first 67%, and then a further 50%, of marine invertebrate species respectively.

temperature increase to the severe greenhouse effect caused by the rapid increase in atmospheric methane and CO₂ following the first pulse of methane release.

The abrupt but less clearly defined shifts in $\delta^{13}\text{C}_{\text{org}}$ above shift C (Fig. 1) could represent further smaller pulses of methane hydrate dissociation. The disappearance of regular $\delta^{13}\text{C}_{\text{org}}$ and CaCO₃ cycles higher in the section (above ~2 m, Fig. 1) prevents the construction of an astronomically calibrated timescale for the entire excursion interval. We attribute the disappearance of the cycles to the likelihood that the climate system would have been in a state of severe flux following large-scale methane hydrate dissociation, and to the consequent effect that this would have had on ocean biogeochemistry, and hence the $\delta^{13}\text{C}_{\text{org}}$ and CaCO₃ signals.

High-precision radiometric dating of the Karoo-Ferrar large igneous province²⁰ has recently highlighted the temporal coincidence between its emplacement and the terrestrial and marine early Toarcian mass extinctions^{16,20}. We suggest that the initiation of long-term global warming in the late Pliensbachian associated with the inception of Karoo-Ferrar volcanic activity^{1,7}, when coupled with solar insolation maxima controlled by precession, was able to periodically exceed a climatic threshold and thereby trigger the thermal dissociation of continental-shelf methane hydrate deposits. Thermal dissociation could have occurred through a steepening of geothermal gradients in continental-shelf sediments owing to bottom water warming²¹, possibly through a change in thermohaline circulation and transfer of bottom water formation from high to low

latitudes⁷. Destabilization and failure of hydrate-bearing sedimentary deposits as a result of thermal dissociation²² is likely to have caused the release of methane over a period of time that was much shorter than that suggested by our <2 kyr estimate for each abrupt shift A, B and C (Fig. 1). The duration of the shifts we observe may instead reflect the much longer time taken for the oceans to mix fully. The most likely cause for the slower, initial ~2‰ decrease in $\delta^{13}\text{C}_{\text{org}}$ that began before the first methane hydrate dissociation pulse (Fig. 1) was effusion of volcanogenic CO₂ ($\delta^{13}\text{C}_{\text{volcanic}} \approx -6\text{‰}$) from the Karoo-Ferrar large igneous province. However, as noted previously⁷, volcanogenic CO₂ could not have effused sufficiently rapidly or in the quantities required to cause the overall $\delta^{13}\text{C}$ excursion.

Our high-resolution analysis of early Toarcian methane hydrate dissociation allows us to compare it directly with another methane hydrate dissociation event that is thought to have occurred at the Palaeocene-Eocene thermal maximum (PETM) ~55 Myr ago^{21,23,24}. High-resolution $\delta^{13}\text{C}$ analysis coupled with a timescale based on astronomical precession through the PETM indicates that up to three pulses of methane release occurred at similar rates and over a period of time comparable to the early Toarcian event^{23,24}, prompting us to speculate that the PETM event was also controlled by astronomical precession. The total mass of carbon calculated to have been released at the PETM is between 1,100 and 2,100 Gt (ref. 21). By contrast, the total mass of carbon released during the early Toarcian event has been calculated to have been ~5,000 Gt (ref. 8). This latter estimate was based on a $\delta^{13}\text{C}$ excursion of ~5‰ in the entire exchangeable carbon reservoir⁸, and may underestimate the true quantity of carbon released owing to the recovery we observe in $\delta^{13}\text{C}_{\text{org}}$ between pulses B and C (Fig. 1). Because we have shown that the dissociation of methane hydrate in the early Toarcian proceeded in three, approximately equal stages, the quantity of methane released from each of these individual pulses was of a mass comparable to the total amount of methane released during the PETM event.

Mechanisms other than methane hydrate dissociation have been proposed for the $\delta^{13}\text{C}$ excursion at the PETM, including global wildfires²⁵, thermogenic methane release from hydrocarbons²⁶, and extraterrestrial comet impact²⁷. However, the greater magnitude of the early Toarcian event, and the fact that the perturbation was astronomically controlled, rules out all of these other mechanisms for the early Toarcian event at least. The only other theory that has gained significant support for causing the early Toarcian $\delta^{13}\text{C}$ excursion is that sinking organic matter remained trapped beneath a stable redox boundary until sudden oceanic overturn, which flooded the surface ocean with ¹²C-rich carbon²⁻⁵. However, simple mass balance arguments demonstrate that each abrupt shift A, B and C would have required the recycling of approximately twice as much organic carbon than is currently present in the modern ocean. In addition, this mechanism cannot account for the coeval increases in global warming and weathering rates observed across the interval, which are the predicted consequences of methane hydrate dissociation and atmospheric CO₂ increase¹.

Our results suggest that the rapid and large-scale dissociation of methane hydrate in the early Toarcian was driven by astronomical forcing superimposed on longer-term global warming, and that these combined events are linked to marine extinction events. Our observations confirm the link between natural processes (volcanism, astronomical cycles) as drivers of abrupt environmental change in the presence of sensitive climatic thresholds. Defining where these thresholds lie may help in the understanding of the impact of anthropogenic climate change.

Received 26 October 2004; accepted 4 July 2005.

1. Cohen, A. S., Coe, A. L., Harding, S. M. & Schwark, L. Osmium isotope evidence for the regulation of atmospheric CO₂ by continental weathering. *Geology* **32**, 157–160 (2004).
2. Küspert, W. G. in *Cyclic and Event Stratification* (eds Einsele, G. & Seilacher, A.) 482–501 (Springer, Berlin, 1982).

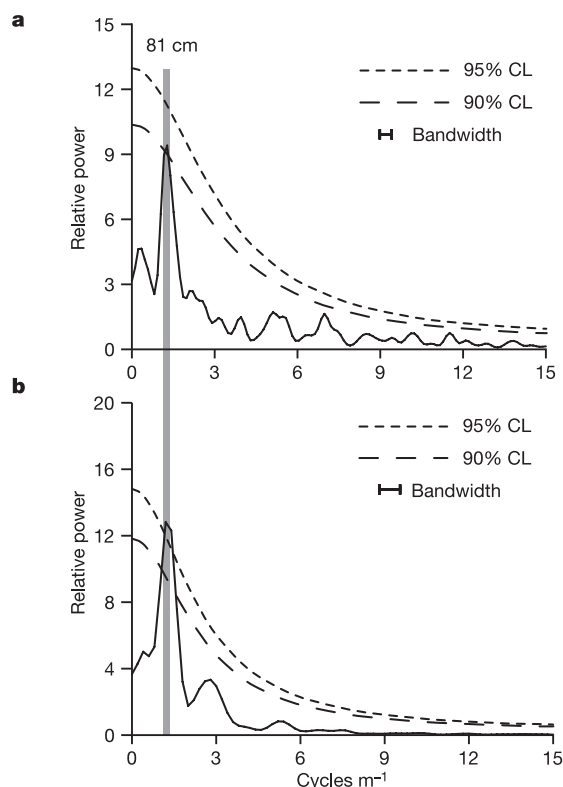


Figure 2 | Spectral analysis of early Toarcian data. **a**, Spectral analysis of CaCO₃ concentration data from -6.34 m to 2.23 m (see Fig. 1 for definition of 0 m). **b**, Spectral analysis of $\delta^{13}\text{C}_{\text{org}}$ data from -4.29 m to 0.70 m. Spectral analysis of $\delta^{13}\text{C}_{\text{org}}$ and CaCO₃ data from the rest of the section (Fig. 1) revealed no regular cycles of statistical significance. Both spectra were constructed using smoothed periodograms computed from the Fourier transform of the raw data. Confidence levels (CL) were obtained using statistical methods presented in ref. 29. A pronounced peak at 1.23 cycles m⁻¹ (81 cm mean cycle wavelength) is present in both spectra above the 90% confidence level. This indicates the presence of a regular cyclicity, which we attribute to 21 kyr astronomical precession.

3. Schouten, S., Van Kaam-Peters, H. M. E., Rijpstra, W. I. C., Schoell, M. & Damste, J. S. Effects of an oceanic anoxic event on the stable carbon isotopic composition of Early Toarcian carbon. *Am. J. Sci.* **300**, 1–22 (2000).
4. Sælen, G., Tyson, R. V., Telnæs, N. & Talbot, M. R. Contrasting watermass conditions during deposition of the Whitby Mudstone (Lower Jurassic) and the Kimmeridge Clay (Upper Jurassic) formations, UK. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **163**, 163–196 (2000).
5. Röhl, H.-J., Schmid-Röhl, A., Oschmann, W., Frimmel, A. & Schwark, L. Erratum to "The Posidonia Shale (Lower Toarcian) of SW-Germany: an oxygen-depleted ecosystem controlled by sea level and palaeoclimate". *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **169**, 273–299 (2001).
6. Jenkyns, H. C., Gröcke, D. R. & Hesselbo, S. P. Nitrogen isotope evidence for water mass denitrification during the early Toarcian (Jurassic) oceanic anoxic event. *Palaeoceanography* **16**, 593–603 (2001).
7. Hesselbo, S. P. *et al.* Massive dissociation of gas hydrate during a Jurassic oceanic anoxic event. *Nature* **406**, 392–395 (2000).
8. Beerling, D. J., Lomas, M. R. & Gröcke, D. R. On the nature of methane gas-hydrate dissociation during the Toarcian and Aptian oceanic anoxic events. *Am. J. Sci.* **302**, 28–49 (2002).
9. Harries, P. J. & Little, C. T. S. The early Toarcian (Early Jurassic) and the Cenomanian-Turonian (Late Cretaceous) mass extinctions: similarities and contrasts. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **154**, 39–66 (1999).
10. Jenkyns, H. C. The Early Toarcian (Jurassic) Anoxic Event - stratigraphic, sedimentary, and geochemical evidence. *Am. J. Sci.* **288**, 101–151 (1988).
11. Jenkyns, H. C., Jones, C. E., Gröcke, D. R., Hesselbo, S. P. & Parkinson, D. N. Chemostratigraphy of the Jurassic System: applications, limitations and implications for palaeoceanography. *J. Geol. Soc. Lond.* **159**, 351–378 (2002).
12. Bailey, T. R., Rosenthal, Y., McArthur, J. M., van de Schootbrugge, B. & Thirlwall, M. F. Paleocceanographic changes of the Late Pliensbachian-Early Toarcian interval: a possible link to the genesis of an Oceanic Anoxic Event. *Earth Planet. Sci. Lett.* **212**, 307–320 (2003).
13. Jenkyns, H. C. Evidence for rapid climate change in the Mesozoic-Palaeogene greenhouse world. *Phil. Trans. R. Soc. Lond. A* **361**, 1885–1916 (2003).
14. Benton, M. J. Diversification and extinction in the history of life. *Science* **268**, 52–58 (1995).
15. Bucefalo Palliani, R., Mattioli, E. & Riding, J. B. The response of marine phytoplankton and sedimentary organic matter to the early Toarcian (Lower Jurassic) oceanic anoxic event in northern England. *Mar. Micropaleontol.* **46**, 223–245 (2002).
16. Pálffy, J., Smith, P. L. & Mortensen, J. K. Dating the end-Triassic and Early Jurassic mass extinctions, correlative large igneous provinces, and isotopic events. *Geol. Soc. Am. Spec. Pap.* **356**, 523–532 (2002).
17. Hinnov, L. A. & Park, J. J. Strategies for assessing Early-Middle (Pliensbachian-Aalenian) Jurassic cyclochronologies. *Phil. Trans. R. Soc. Lond. A* **357**, 1831–1859 (1999).
18. Hallam, A. Estimates of the amount and rate of sea-level change across the Rhaetian-Hettangian and Pliensbachian-Toarcian boundaries (latest Triassic to early Jurassic). *J. Geol. Soc. Lond.* **154**, 773–779 (1997).
19. Cope, J. C. W. Discussion on estimates of the amount and rate of sea-level change across the Rhaetian-Hettangian and Pliensbachian-Toarcian boundaries (latest Triassic to Early Jurassic). *J. Geol. Soc. Lond.* **155**, 421–422 (1998).
20. Duncan, R. A., Hooper, P. R., Rehacek, J., Marsh, J. S. & Duncan, A. R. The timing and duration of the Karoo igneous event, southern Gondwana. *J. Geophys. Res.* **102**, 18127–18138 (1997).
21. Dickens, G. R., O'Neil, J. R., Rea, D. K. & Owen, R. M. Dissociation of oceanic methane hydrate as a cause of the carbon isotope excursion at the end of the Paleocene. *Paleoceanography* **10**, 965–971 (1995).
22. Paull, C. K., Buelow, W. J., Ussler, W. & Borowski, W. S. Increased continental-margin slumping frequency during sea-level lowstands above gas hydrate-bearing sediments. *Geology* **24**, 143–146 (1996).
23. Bains, S., Corfield, R. M. & Norris, R. D. Mechanisms of climate warming at the end of the Paleocene. *Science* **285**, 724–727 (1999).
24. Röhl, U., Bralower, T. J., Norris, R. D. & Wefer, G. New chronology for the late Paleocene thermal maximum and its environmental implications. *Geology* **28**, 927–930 (2000).
25. Kurtz, A. C., Kump, L. R., Arthur, M. A., Zachos, J. C. & Paytan, A. Early Cenozoic decoupling of the global carbon and sulphur cycles. *Paleoceanography* **18**, 1–14 (2003).
26. Svensen, H. *et al.* Release of methane from a volcanic basin as a mechanism for initial Eocene global warming. *Nature* **429**, 542–545 (2004).
27. Kent, D. V. *et al.* A case for a comet impact trigger for the Paleocene/Eocene thermal maximum and carbon isotope excursion. *Earth Planet. Sci. Lett.* **211**, 13–26 (2003).
28. Howarth, M. K. The ammonite family Hildoceratidae in the lower Jurassic of Britain. *Palaeogr. Soc. Monogr.* **145**, 1–200 (1992).
29. Mann, M. E. & Lees, J. M. Robust estimation of background noise and signal detection in climatic time series. *Clim. Change* **33**, 409–445 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements D.B.K. was supported by a NERC CASE studentship with Bartington Instruments Ltd. We thank colleagues in the Department of Earth Sciences, The Open University, and the Geologisches Institut, Universität zu Köln, for analytical assistance and comments on an earlier draft of this manuscript.

Author Contributions D.B.K., A.L.C. and A.S.C. collected samples and contributed equally to interpretation. L.S. was responsible for the carbon-isotope analyses.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.B.K. (d.kemp@open.ac.uk).

LETTERS

River plumes as a source of large-amplitude internal waves in the coastal ocean

Jonathan D. Nash¹ & James N. Moum¹

Satellite images have long revealed the surface expression of large amplitude internal waves that propagate along density interfaces beneath the sea surface^{1–3}. Internal waves are typically the most energetic high-frequency events in the coastal ocean^{4–6}, displacing water parcels by up to 100 m and generating strong currents and turbulence⁷ that mix nutrients into near-surface waters for biological utilization. While internal waves are known to be generated by tidal currents over ocean-bottom topography^{8–13}, they have also been observed frequently in the absence of any apparent tide-topography interactions^{1,7,14}. Here we present repeated measurements of velocity, density and acoustic backscatter across the Columbia River plume front. These show how internal waves can be generated from a river plume that flows as a gravity current into the coastal ocean. We find that the convergence of horizontal velocities at the plume front causes frontal growth and subsequent displacement downward of near-surface waters. Individual freely propagating waves are released from the river plume front when the front's propagation speed decreases below the wave speed in the water ahead of it. This mechanism generates internal waves of similar amplitude and steepness as internal waves from tide-topography interactions observed elsewhere¹¹, and is therefore important to the understanding of coastal ocean mixing.

It is generally assumed that internal waves radiate from locations where tidal currents flow over topographic features such as shelf-breaks¹⁰, banks¹¹ and sills^{12,13}. In the last case, waves formed downstream of a sill are trapped to the topography when their wavespeed c equals that of the opposing tidal flow u . They are released and propagate upstream as free waves when u slackens below c (refs 8, 9, 15). The Froude number $F = u/c \leq 1$ sets the criterion for free wave propagation.

In the atmosphere, gravity currents¹⁶ are well-known to excite large-amplitude waves. Perhaps the most famous is the 'Morning Glory', a series of ~500-m amplitude undulations over the Gulf of Carpentaria off northern Australia^{17,18}. Wave generation from gravity currents has also been observed in thunderstorm outflows¹⁹ and mountain slope drainage winds²⁰. However, the large scales of atmospheric flows make it difficult to obtain the detailed measurements necessary to show the process by which freely propagating waves emerge from a gravity current. Although laboratory experiments^{21,22} have helped to show this evolution, these experiments were limited to small, sub-geophysical scales. Neither atmospheric nor laboratory observations have clearly defined the criterion for wave release.

Rivers issue into the coastal ocean as tidally modulated pulses of fresh water that form positively buoyant gravity currents²³. The evolving properties of these gravity currents are determined by the initial momentum at the river's mouth, by interactions with coastal currents and winds, and by the Earth's rotation, which tends to turn the current to the right in the Northern Hemisphere. These factors all

affect the location, propagation speed and sharpness of the gravity current front.

Satellite images capture single snapshots of waves radiating from the mouth of the Columbia River (Fig. 1; refs 1, 2). However, they provide no information on the waves' internal structure. Nor do they show the sequence of events leading to their generation, since these images are acquired infrequently (<1 per day). Our recent *in situ* observations across a front at the northern edge of the tidally pulsing Columbia River plume provide the necessary sequencing to clearly define the condition for the formation of large-amplitude internal waves from a gravity current. By analogy to topographic release of waves from a sill, this condition is described in terms of a Froude number.

Satellite sea surface temperature (SST) distinguishes the warm, summertime plume from cold, recently upwelled coastal waters (Fig. 2a–c) on 23 July 2004. High tide coincided with the image shown in Fig. 2a. Plume remnants from the previous tidal cycle's discharge appear in the offshore thermal structure. Ebb currents started to flow

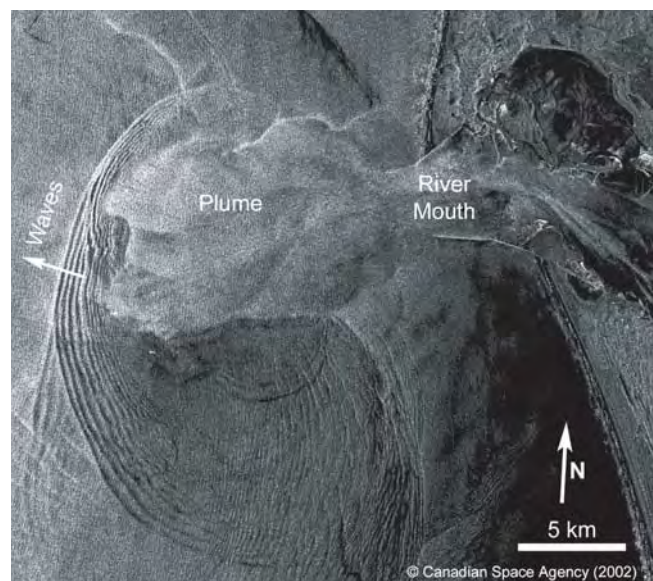


Figure 1 | Synthetic aperture radar (SAR) image of the Columbia River plume on 9 August 2002. Image indicates regions of enhanced surface roughness associated with plume-front and internal wave velocity convergences. Similar features appear in images during all summertime months (April–October; see <http://oceanweb.ocean.washington.edu/rise/data.htm> for more Columbia River plume images) and from other regions^{1,2}. SAR image courtesy of P. Orton, T. Sanders and D. Jay; image was processed at the Alaska Satellite Facility, and is copyright Canadian Space Agency.

¹College of Oceanic and Atmospheric Sciences, Oregon State University, 104 COAS Admin Bldg, Oregon State University, Corvallis, Oregon 97331, USA.

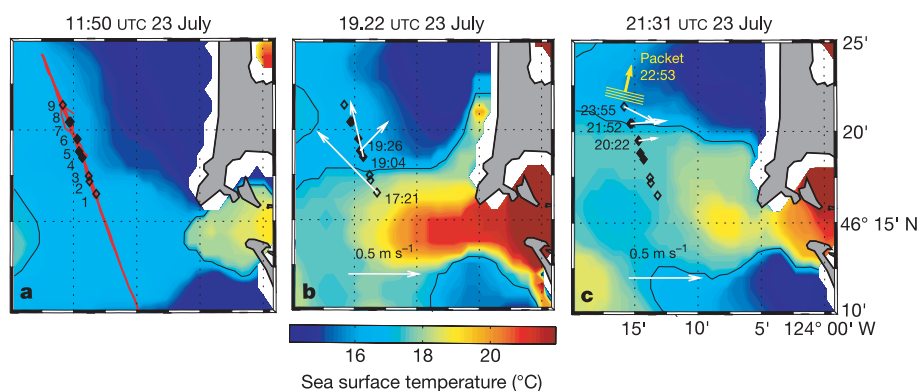


Figure 2 | Progression of the Columbia River plume from satellite-derived SST images. Times (23 July 2004 UTC) are 11:50 (a), 19:22 (b) and 21:31 (c). Red line in left panel is ship track. Diamonds show locations where plume front was crossed; filled diamonds correspond to the four crossings presented in Fig. 3. Near-surface fluid velocities behind the plume front at

selected crossings are indicated in b and c (45-s average over $0\text{ m} < z < 5\text{ m}$); vectors are grouped to correspond to time of SST images. The 17°C isotherm is contoured and represents the approximate front location. Location of the wave packet at 22:53 as imaged by shipboard X-band radar is shown in c.

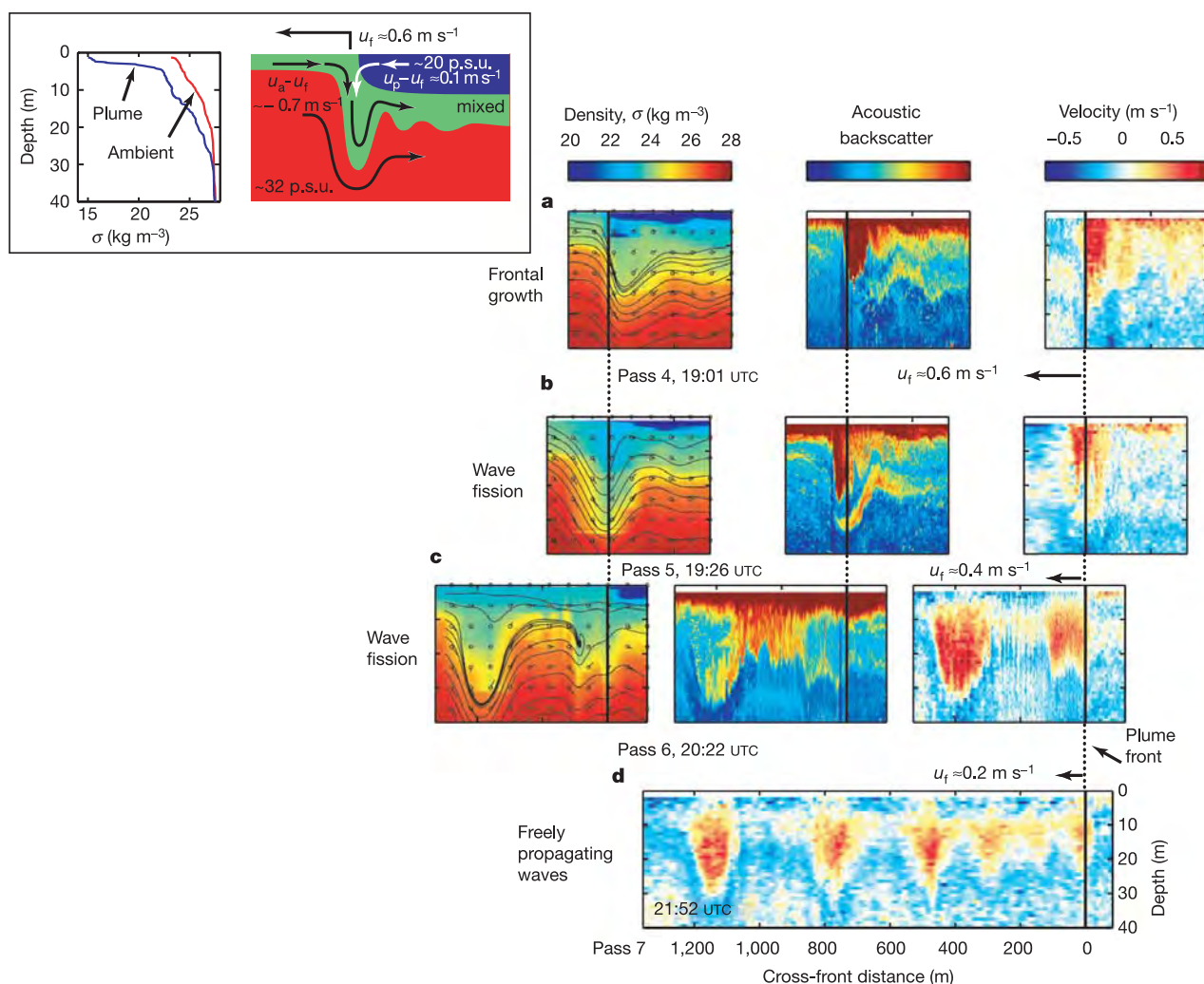


Figure 3 | Three stages of a wave-generation event. a, Frontal growth at the plume's leading edge; b, c, wave fission from the plume front; and d, free propagation of a train of large-amplitude internal waves. Shown are density (left), acoustic backscatter (middle) and cross-front component of horizontal velocity (right) in a reference frame aligned with the front (see Methods). Positive distances and velocities are approximately northward; velocities are relative to a stationary reference frame. Panels are shifted to align the plume front (as determined from sea surface salinity); vertical black lines represent the plume front (zero cross-front distance). Only cross-front

velocity is shown for the freely propagating waves in d. Particle streamlines and velocity vectors (u, w) in a reference frame moving with the front (translating at speed u_f as indicated) are contoured over the density plots. A schematic cartoon illustrating frontal growth in a reference frame moving with the plume front (at speed u_f) is shown in upper left inset. Velocities of the near-surface fluid behind the plume front (u_p) and ambient water ahead of it (u_a) are indicated. Also shown for pass 4 (panel a) are vertical profiles of density ahead of (red, ambient) and behind (blue, plume) the front.

shortly thereafter, producing a pulse of fresh and warm water visible in Fig. 2b. At 19:00 UTC, the plume front flowed northward as a gravity current with $u_f \approx 0.6 \text{ m s}^{-1}$ (Fig. 2b), opposed by weak winds from the north and a southward near-surface current ($u_a \approx -0.1 \text{ m s}^{-1}$) that helped sharpen the front. (u_f and u_a represent the cross-front components of the frontal propagation speed and the velocity of ambient fluid ahead of the front.) 1.5 h later, plume-front velocities turned eastward (Fig. 2c). During this transition period, a packet of waves was released.

Cross-front snapshots of velocity, density and acoustic backscatter capture the growth of the plume front, the release of two individual waves, and the free propagation of a train of waves away from the plume front (Fig. 3). In its initial phase, the plume front moved northward at $\sim 0.6 \text{ m s}^{-1}$ and vertical displacements at the front grew to 20 m during a ~ 2 -h period (Fig. 3a). Twenty-five minutes later (Fig. 3b), the cross-front component of the plume velocity weakened and the isopycnal depression at the plume front (now almost 25 m) separated from it. One hour later (Fig. 3c), this distinct and solitary wave had propagated 400 m ahead of the plume front, and a second wave emerged. Following this, a train of freely propagating internal waves were observed (Fig. 3d). The full generation sequence is summarized as a progression: frontal growth $\rightarrow$ wave fission $\rightarrow$ freely propagating waves. The space-time plot (Fig. 4) of front, wave and ship locations illustrates this progression.

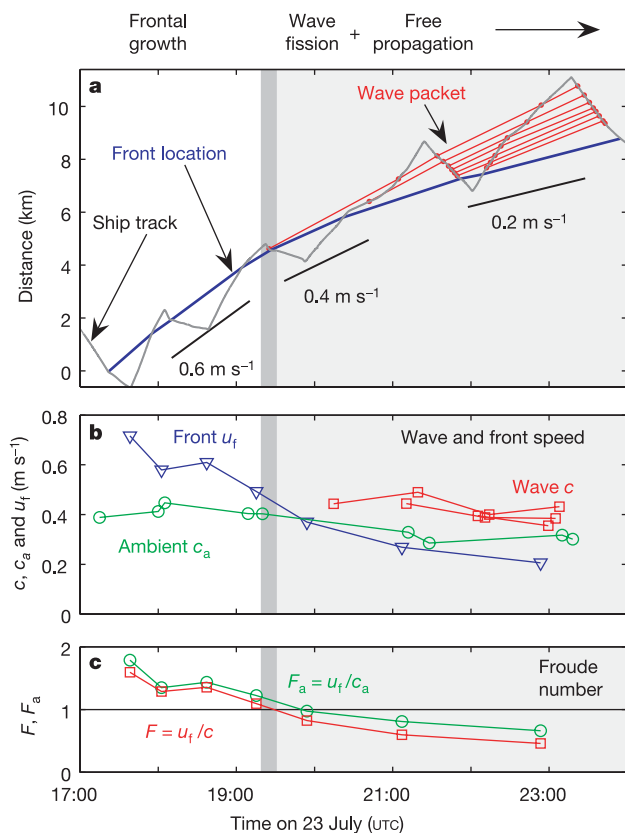


Figure 4 | Time-evolution of plume front and wave packet. **a**, Location of ship (grey), plume front (blue) and wavepacket (red) as a function of time. Ordinate is distance from the first frontal crossing along a curved trajectory perpendicular to either plume front or leading wave. **b**, The speed of the plume front (u_f , blue) as computed from the rate of change of front location. The first-mode linear wavespeed ahead of the plume (c_a , green) reflects the slowly-changing background density and velocity structure in the ambient ocean; observed wavespeeds computed from first-differences of wave location (c , red) are higher. Corresponding Froude numbers are shown in **c**; F was computed using $c = 0.45 \text{ m s}^{-1}$. The vertical grey bar represents the first wave fission event (Fig. 3b) and $F = 1$; light shading represents the domain over which fission and free propagation are permitted ($F < 1$).

In analogy to topographic control, the Froude number $F = u_f/c$ based on the frontal velocity u_f is a natural parameter governing the flow^{22,24,25}. In a reference frame moving with the front, F represents the ratio of the opposing flow speed required to keep the front stationary (u_f), to the internal wavespeed in the medium into which the front advances (c).

In a fluid with vertical gradients in both density and velocity, the intrinsic speed of a long, small-amplitude linear wave is determined from the hydrostatic Taylor-Goldstein equation²⁶. Intrinsic wavespeeds in the ambient coastal waters ($c_a \approx 0.4 \text{ m s}^{-1}$; Fig. 4b) are roughly half that within the highly stratified brackish plume. Measured wavespeeds in the ambient waters (c) are $\sim 30\%$ greater than c_a owing to finite amplitude effects⁶. We may thus form two Froude numbers: (1) $F_a = u_f/c_a$ based on the intrinsic properties of the ambient fluid, and (2) $F = u_f/c$ based on the measured wavespeed in that medium (Fig. 4b, c). F_a always exceeds F .

Initially the front moves at a speed u_f that exceeds both c_a and c ($F > 1$; Fig. 4b). During this phase, horizontal velocity convergence at the plume's edge is intense (Fig. 3a; $\Delta u \approx 1 \text{ m s}^{-1}$ in $\Delta x' = 10 \text{ m}$) and drives $> 0.3 \text{ m s}^{-1}$ vertical velocities. (See Methods for definition of x' .) This convergence provides a means of converting the gravity current's kinetic energy to potential energy. Since the frontal velocity exceeds all internal wavespeeds ahead of the front, this potential energy is trapped at the front and cannot radiate ahead of it (Fig. 4a, b). Waves may, however, propagate back towards the plume source²⁷, but these were not captured in this study.

As the plume front decelerates, F decreases below unity (Fig. 4c); wave fission ensues. With the transition of F from super- ($F > 1$) to subcritical ($F < 1$), the depression that was originally locked to the freshwater front advances into the ambient fluid as a freely propagating wave. In this way, the wave inherits the vertical displacement structure of the front. Because c increases with amplitude⁶, F also decreases as frontal vertical displacements grow. It is therefore impossible to predict the precise timing of fission from F_a alone. Only when frontal amplitudes are sufficiently large does c exceed u_f , permitting fission to occur.

Following fission of the first wave (Fig. 3b), sustained convergence at the front continues to displace fluid downward, creating anew the disturbance from which subsequent waves emerge. In Fig. 3b, the wave is $\sim 50 \text{ m}$ ahead of the freshwater front, moving to 400 m ahead in Fig. 3c, by which time the front has regenerated its vertical displacement to sufficient amplitude for a second wave to release.

Once released, waves propagate freely at $c \approx 0.40 - 0.45 \text{ m s}^{-1}$ with 20-m amplitude (Fig. 4b). These were tracked more than 5 km from their release location (Fig. 4a). Meanwhile, the increasingly subcritical front forces waves with successively smaller amplitude. This factor, together with reduced convergence, limits frontal growth and restricts the amplitudes of the released waves. By the time of our last crossing, the frontal velocity gradient had been reduced to $< 0.4 \text{ m s}^{-1}$ over 1,000 m. Ultimately, the front loses its velocity signature entirely.

In summary, internal waves generated from the Columbia River plume are of similar amplitude and steepness to those generated over topography elsewhere in the coastal ocean¹⁰. Although less energetic than some waves which propagate through deep water (for example, through the South China Sea²⁸), these plume-generated internal waves are large compared to the local water depth, and have important implications for biology and turbulent mixing.

Wave fission from a decelerating gravity current represents an important mechanism for generation of large-amplitude internal waves in the coastal ocean and explains their existence in the absence of a topographic generation site^{7,14}. The Froude number criterion controlling the timing of wave fission is analogous to that of topographic generation; that is, in each case, free propagation occurs when the wavespeed exceeds the background velocity that arrests the disturbance. For topographic control, that velocity is relative to topography⁸; for a gravity current, that velocity is relative to the

propagating front²². This mechanism will be realized for any river with discharge velocities exceeding coastal internal wavespeeds.

METHODS

Density, biological fluorescence and turbulence profiles were obtained from within 2 m of the surface to the bottom using the Chameleon turbulence profiler²⁹; its horizontal resolution is limited by the unequal but nominally 100-m spacing between profiles. Our perspective of the structure of the waves and front is augmented with a rapidly sampled echosounder (Biosonics 120 kHz; acoustic scatterers include zooplankton and density microstructure) and acoustic Doppler current profiler (ADCP; RD Instruments 300 kHz), both mounted 1 m beneath the sea surface.

Nine transects across the plume front were acquired as part of an interdisciplinary effort to understand river influences in coastal ecosystems (<http://www.ocean.washington.edu/rise/index.htm>). Front and wave locations were determined from Chameleon density and ADCP velocity profiles. Frontal orientation was determined by combining X-band shipboard radar and ADCP velocity. The front was assumed perpendicular to the wave-induced fluid velocities, consistent with available radar images. The distance from the leading wave x' is $x' = [\mathbf{x} - \mathbf{x}_w(t)] \cdot \mathbf{n}$, where $\mathbf{x}$ is a measurement location, $\mathbf{x}_w(t)$ is the location of the leading wave at time t as computed from a linear interpolation of the crossings, and $\mathbf{n}$ is the unit vector normal to the front (and in the direction of wave propagation). The distance between the leading wave and the plume front (as determined from the surface salinity) is subtracted from x' to yield the cross-front distance in Fig. 3. This coordinate transformation minimizes spatial Doppler shifting (the tendency for features to look compressed or elongated when measured from a moving platform) at the expense of smearing temporal information (approximately 1 h of data goes into each image). We implicitly assume that changes in the waveform are small over each transect's duration.

Eigenvalues and eigenvectors of the Taylor-Goldstein equation provide the propagation speeds and vertical structure functions for long wavelength, small amplitude, hydrostatic disturbances²⁶. These solutions depend only on the density stratification and velocity shear of the background flow, so they are intrinsic to the medium. To predict finite amplitude wavespeeds, higher-order corrections are required, such as that provided through solutions of the Korteweg de Vries (KdV) equation⁶. Solutions to the KdV equation for a 20-m wave in a non-sheared medium suggest an increase of 30–40% in wavespeed over the linear modes (when computed using the Taylor-Goldstein equation in a non-sheared medium). This is in accord with measured wavespeeds, for which a similar ~30% increase is observed as compared to the linear wavespeeds in a sheared medium. Finite amplitude wavespeeds in a sheared medium have not been computed.

Received 10 March; accepted 13 June 2005.

1. Fu, L. L. & Holt, B. *Seasat Views Oceans and Sea Ice with Synthetic-Aperture Radar* (JPL publication 81-120, NASA Jet Propulsion Laboratory, Pasadena, 1982).
2. Jackson, C. & Apel, J. An atlas of internal solitary-like waves and their properties. (http://www.internalwaveatlas.com/Atlas2_index.html) 2004.
3. Ray, R. D. & Mitchum, G. T. Surface manifestation of internal tides generated near Hawaii. *Geophys. Res. Lett.* **23**, 2101–2104 (1996).
4. Huthnance, J. M. Internal tides and waves near the continental shelf edge. *Geophys. Astrophys. Fluid Dyn.* **48**, 81–105 (1989).
5. Osborne, A. R. & Burch, T. L. Internal solitons in the Andaman Sea. *Science* **208**, 451–460 (1980).
6. Ostrovsky, L. & Stepanyants, Y. Do internal solitons exist in the ocean? *Rev. Geophys.* **27**, 293–310 (1989).
7. Moun, J. N., Farmer, D. M., Smyth, W. D., Armi, L. & Vagle, S. Structure and

generation of turbulence at interfaces strained by internal solitary waves propagating shoreward over the continental shelf. *J. Phys. Oceanogr.* **33**, 2093–2112 (2003).

8. Maxworthy, T. A note on the internal solitary waves produced by tidal flow over a three-dimensional ridge. *J. Geophys. Res.* **84**, 338–346 (1979).
9. Lamb, K. G. Numerical experiments of internal wave generation by strong tidal flow across a finite amplitude bank edge. *J. Geophys. Res.* **99**, 843–864 (1994).
10. Chereskin, T. K. Generation of internal waves in Massachusetts Bay. *J. Phys. Oceanogr.* **88**, 2649–2661 (1983).
11. Loder, J. W., Brickman, D. & Horne, E. P. W. Detailed structure of currents and hydrography on the northern side of Georges Bank. *J. Geophys. Res.* **97**, 14331–14351 (1992).
12. Farmer, D. M. & Smith, J. D. in *Hydrodynamics of Estuaries and Fjords* (ed. Nihoul, J.) 465–493 (Elsevier, Amsterdam, 1978).
13. Farmer, D. M. & Armi, L. The generation and trapping of solitary waves over topography. *Science* **283**, 188–190 (1999).
14. Stanton, T. P. & Ostrovsky, L. A. Observations of highly nonlinear internal solitons over the continental shelf. *Geophys. Res. Lett.* **25**, 2695–2698 (1998).
15. Sutherland, B. R. Interfacial gravity currents. I. Mixing and entrainment. *Phys. Fluids* **14**, 2244–2254 (2002).
16. Simpson, J. E. *Gravity Currents in the Environment and Laboratory* (Cambridge Univ. Press, Cambridge, UK, 1987).
17. Christie, D. R., Muirhead, K. J. & Clarke, R. H. Solitary waves in the lower atmosphere. *Nature* **293**, 46–49 (1981).
18. Smith, R. K., Crook, N. & Roff, G. The Morning Glory: An extraordinary atmospheric undular bore. *Q. J. R. Meteorol. Soc.* **108**, 937–956 (1982).
19. Doviak, R. J., Chen, S. S. & Christie, D. R. Thunderstorm-generated solitary wave observation compared with theory for nonlinear waves in a sheared atmosphere. *J. Atmos. Sci.* **48**, 87–111 (1991).
20. Rao, M. P., Castracane, P., Casadio, S., Fua, D. & Fiocco, G. Observations of atmospheric solitary waves in the urban boundary layer. *Boundary-Layer Meteorol.* **111**, 85–108 (2004).
21. Rottman, J. W. & Simpson, J. E. The formation of internal bores in the atmosphere: A laboratory model. *Q. J. R. Meteorol. Soc.* **115**, 941–963 (1989).
22. Maxworthy, T., Leilich, J., Simpson, J. E. & Meiburg, E. H. The propagation of a gravity wave into a linearly stratified fluid. *J. Fluid Mech.* **453**, 371–394 (2002).
23. Orton, P. M. & Jay, D. A. Observations at the tidal plume front of a high-volume river outflow. *Geophys. Res. Lett.* **32**, L11605 doi:10.1029/2005GL02237 (2005).
24. Britter, R. E. & Simpson, J. E. Experiments on the dynamics of a gravity current head. *J. Fluid Mech.* **88**, 223–240 (1978).
25. Luketina, D. A. & Imberger, J. Characteristics of a surface buoyant jet. *J. Geophys. Res.* **92**, 5435–5447 (1987).
26. Drazin, P. G. & Reid, W. H. *Hydrodynamic Stability* (Cambridge Univ. Press, Cambridge, 1981).
27. O'Donnell, J. & Garvine, R. W. A time dependant, two layer frontal model of buoyant plume dynamics. *Tellus A* **35**, 73–80 (1983).
28. Duda, T. F. et al. Internal tide and nonlinear internal wave behaviour at the continental slope in the Northern South China Sea. *IEEE J. Ocean. Eng.* **20**, 1105–1130 (2004).
29. Moun, J. N., Gregg, M. C., Lien, R. C. & Carr, M. Comparison of turbulence kinetic energy dissipation rate estimates from two ocean microstructure profilers. *J. Atmos. Ocean. Technol.* **12**, 346–366 (1995).

Acknowledgements We thank M. Neeley-Brown, R. Kreth and A. Perlin for their technical expertise. L. Kilcher, T. Kimura, R. Bjorkquist, A. Horner-Devine, T. Chisholm, and the captain and crew of the RV *Pt. Sur* made data collection possible. Satellite imagery was provided by P.T. Strub and P. Orton. Comments were provided by W.D. Smyth, G. Avicola and J. Klymak. This work was funded by the National Science Foundation and the Office of Naval Research.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.D.N. (nash@coas.oregonstate.edu).

LETTERS

Predecessors of the giant 1960 Chile earthquake

Marco Cisternas¹, Brian F. Atwater², Fernando Torrejón³, Yuki Sawai⁵, Gonzalo Machuca⁴, Marcelo Lagos⁶, Annaliese Eipert⁷, Cristián Youlton¹, Ignacio Salgado¹, Takanobu Kamataki⁵, Masanobu Shishikura⁵, C. P. Rajendran⁸, Javed K. Malik⁹, Yan Rizal¹⁰ & Muhammad Husni¹¹

It is commonly thought that the longer the time since last earthquake, the larger the next earthquake's slip will be. But this logical predictor of earthquake size¹, unsuccessful for large earthquakes on a strike-slip fault², fails also with the giant 1960 Chile earthquake of magnitude 9.5 (ref. 3). Although the time since the preceding earthquake spanned 123 years (refs 4, 5), the estimated slip in 1960, which occurred on a fault between the Nazca and South American tectonic plates, equalled 250–350 years' worth of the plate motion^{3,6–10}. Thus the average interval between such giant earthquakes on this fault should span several centuries^{3,9,10}. Here we present evidence that such long intervals were indeed typical of the last two millennia. We use buried soils and sand layers as records of tectonic subsidence and tsunami inundation at an estuary midway along the 1960 rupture. In these records, the 1960 earthquake ended a recurrence interval that had begun almost four centuries before, with an earthquake documented by Spanish conquistadors in 1575. Two later earthquakes, in 1737 and 1837, produced little if any subsidence or tsunami at the estuary and they therefore probably left the fault partly loaded with accumulated plate motion that the 1960 earthquake then expended.

The 1960 Chile mainshock resulted from a rupture nearly 1,000 km long on a north–south trending fault that conveys the subducting Nazca plate beneath South America at rates averaging 8 m per century³. Lurching westward above the rupture, the South America plate rose in a mostly offshore area while subsiding 1–2 m in a coastal downwarp⁶ (Fig. 1b). The ensuing tsunami, with crests 10–15 m high in Chile¹¹, reached maximum heights of 10 m in Hawaii¹² and 6 m in Japan¹³.

The 1960 earthquake was preceded historically by earthquakes in 1575, 1737 and 1837 (Fig. 1b; Supplementary Table S1). The reported effects from 1575 most nearly resemble those from 1960 (ref. 4). Conquistadors, at forts limited to the northern half of the 1960 rupture area, wrote of persistent marine inundation near Imperial, Valdivia and Castro that implies widespread tectonic subsidence. They also described a devastating tsunami near Valdivia (Supplementary Table S1, record 1). The 1737 earthquake, known only from secondary sources, damaged the few Spanish settlements then remaining south of Concepción. It lacks a reported tsunami, even though tsunamis from central Chile in 1730 and 1751 were noted locally¹⁴ and in Japan^{13,15}. The 1837 earthquake damaged towns along the central third of the 1960 rupture area and changed land levels along the southern half of that area. Its associated tsunami, by reportedly cresting 6 m high in Hawaii¹², provides evidence that the 1837 earthquake released almost half the seismic moment of the

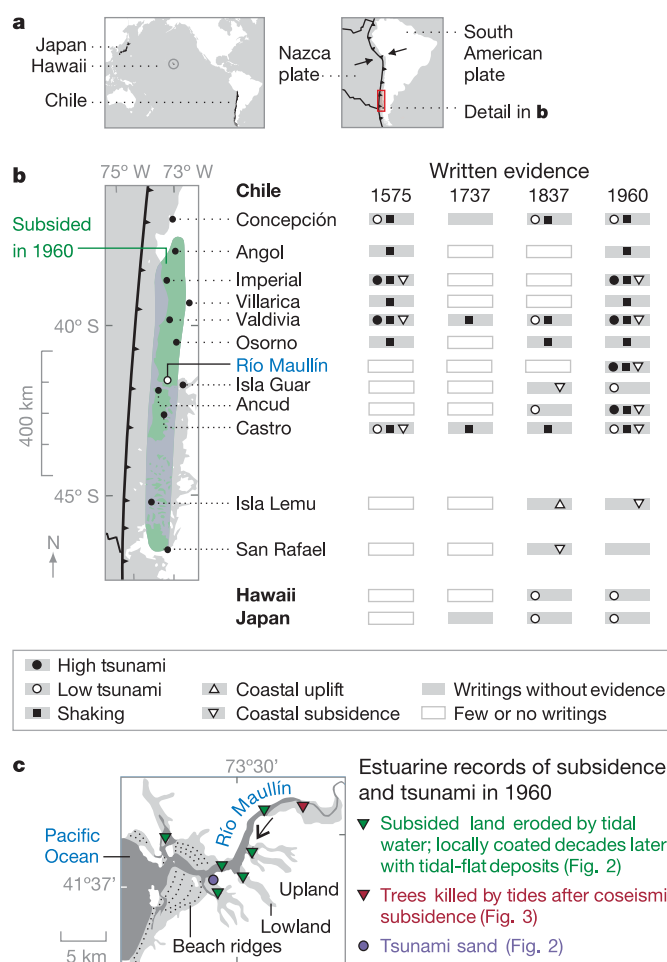


Figure 1 | Index maps. a, Plate-tectonic setting of south-central Chile. Paired arrows indicate plate convergence at 8.4 cm yr⁻¹. **b**, Documented effects of the 1960 earthquake and its historical predecessors. Compiled from refs 4, 5, 13 and 14, and from Supplementary Table S1. **c**, Study area along the Río Maullín. Barbed lines in **a** and **b** show seaward edges of subduction zones; teeth point down the plate boundary.

¹Facultad de Agronomía, Pontificia Universidad Católica de Valparaíso, Casilla 4-D, Quillota, Chile. ²US Geological Survey at University of Washington, Seattle, Washington 98195-1310 USA. ³Centro EULA-Chile. ⁴Departamento de Ciencias Históricas y Sociales, Universidad de Concepción, Casilla 160-C, Concepción, Chile. ⁵Active Fault Research Center, Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology, Tsukuba 305-8567 Japan. ⁶Instituto de Geografía, Pontificia Universidad Católica de Chile, Casilla 306, Santiago, Chile. ⁷National Oceanic and Atmospheric Administration, Seattle, Washington 98115-6349, USA. ⁸Centre for Earth Science Studies, Thiruvananthapuram 69603, India. ⁹Civil Engineering, Indian Institute of Technology, Kanpur 208016, India. ¹⁰Geological Department, Institute of Technology, Bandung 40132, Indonesia. ¹¹Potential Geophysics Division, Meteorological and Geophysical Agency of Indonesia, Jakarta 10720, Indonesia.

1960 mainshock¹⁶. However, according to primary sources in Supplementary Table S1, the same tsunami caused little if any flooding at Valdivia and no reported damage anywhere in Chile or Ancud.

To further compare the 1960 earthquake with these historical earthquakes, and to gain perspective from earlier earthquakes as well, we reconstructed a 2,000-yr history of repeated subsidence and tsunamis at the Río Maullín estuary (Fig. 1b, c). Because of the estuary's central location, this history probably includes earthquakes from full-length breaks of the 1960 rupture area, while perhaps excluding earthquakes from partial ruptures to the north or south.

Our stratigraphic records are tied to modern analogues from 1960

along a nearly marine reach of the Río Maullín. There, 8 km inland from the sea (Fig. 1c, purple dot), markers of the 1960 earthquake extend across faint terraces and beach ridges stranded by net late Holocene emergence¹⁷. Eyewitnesses recall that the 1960 tsunami coated upper terraces with sand¹⁸. We traced the sand, up to 15 cm thick, more than 1 km inland across the buried 1960 soil in areas covered only by the highest post-earthquake tides (Figs 2a, b). In this same area, the sandy record of post-1960 storms extends just a few metres inland from the shore. On lower terraces, now covered routinely by tides, a 1960 pasture soil has been eroded and bioturbated on post-earthquake tidal flats. Waves and currents are now

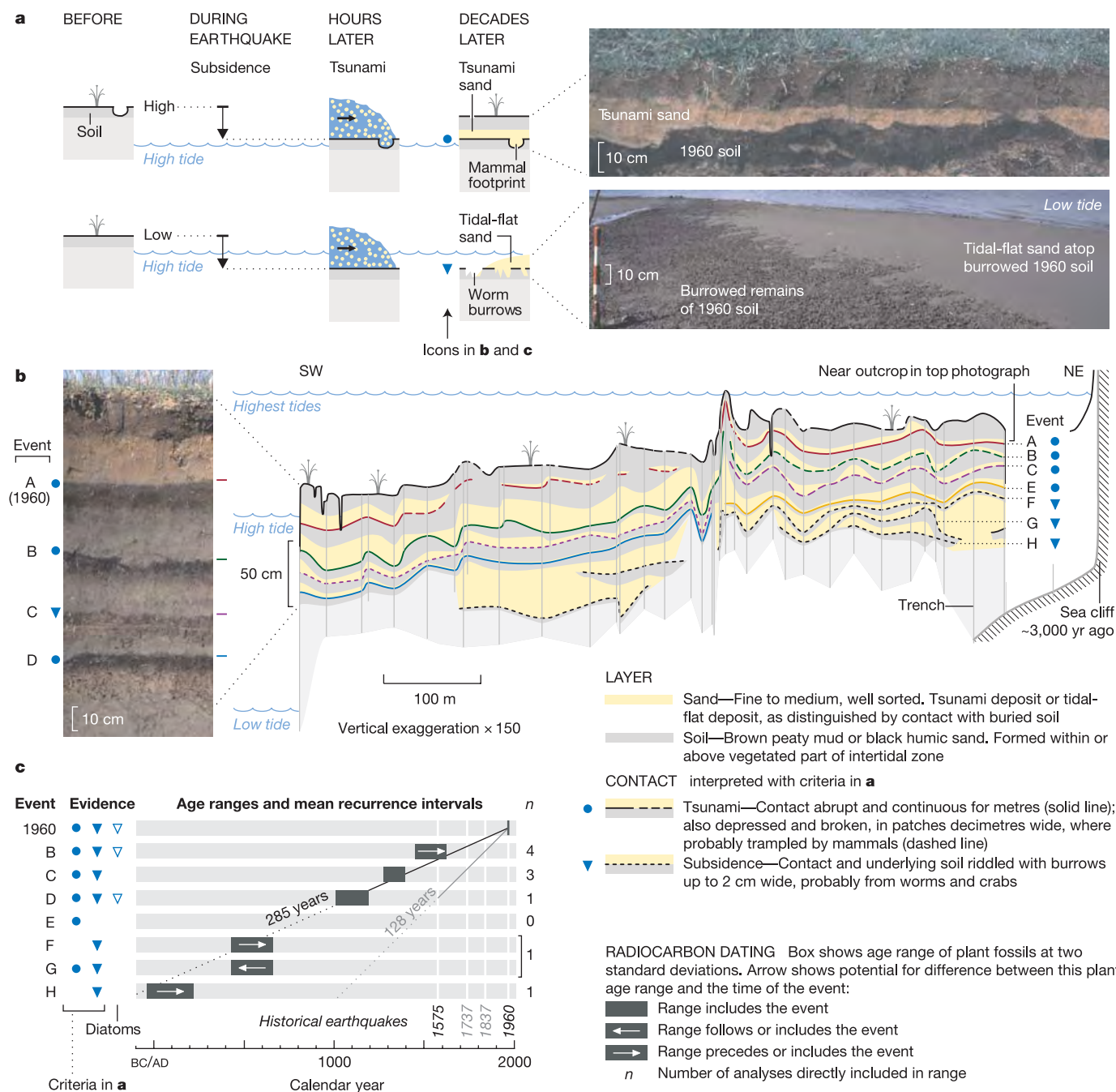


Figure 2 | Stratigraphic evidence for 1960 earthquake and its ancestors in area shown by purple dot in Fig. 1c. Supporting data in Supplementary Figs S1–S4 and Supplementary Tables S2–S4. **a**, Records of the 1960 earthquake that serve as modern analogues for inferring past occurrence of a tsunami and of coseismic subsidence. **b**, Sequences of such records correlated among trenches. Tides measured 1989, 2003, and 2004.

c, Chronology of the inferred events compared with the historical sequence in Fig. 1b. Field evidence for subsidence and tsunami (solid blue circles and triangles at left) comes from all transects (Supplementary Figs S2 and S3). The average of the historical recurrence intervals, 128 yr, contrasts with the longer average intervals between the events recorded stratigraphically.

burying the remains of this soil with sand as much as 1 m thick (Fig. 2a), and with mud in sheltered areas.

Additional sand sheets mantle buried marsh and meadow soils beneath the 1960 soil. Using criteria from the 1960 examples, we interpret some of these sand sheets as tsunami deposits (blue dots in Fig. 2) and others as indicators of subsided, post-earthquake tidal flats (blue triangles). We traced these event records, which probably represent eight earthquakes in all (events A–H), among 60 trenches scattered along 2 km of transects (example, Fig. 2b). Like the 1960 earthquake (event A), four earlier events (B–D, G) produced tsunami deposits on meadows that post-earthquake tides rarely reached and correlative tidal-flat deposits on lower ground. Such evidence, assembled from all transects, is summarized by solid blue symbols in Fig. 2c. Some events are recorded less widely than the 1960 earthquake. The D sand sheet tapers landward without crossing a former beach ridge. The E sand sheet, found entirely inland from that ridge, may have been removed by erosion on the seaward side.

Diatom assemblages from soils that shortly predate and postdate tsunami deposition provide further evidence for subsidence during events A, B and D. In all three cases the assemblages above the tsunami sand are more nearly marine than those in the soil below (summary, Fig. 2c). The difference is clearest for the 1960 event. An attempted comparison for event C failed because the upper part of the buried soil is probably missing from erosion on a post-C tidal flat, and because the remnant soil is contaminated with burrow-filling tidal-flat sand.

In sum, our stratigraphy and paleoecology provide evidence for seven inferred pre-1960 earthquakes from the past 2,000 years (Fig. 2c). The youngest three (B–D), each marked by evidence for both subsidence and tsunami, occurred within the past 1,000 years. Event D dates to the two-sigma range AD 1020–1180—the age of growth-position stem bases of a rush (*Juncus procerus*) that tsunami sand surrounded. The event C tsunami similarly left sand around *Juncus balticus* and *Scirpus americanus* culms in a swale along a spur transect (Supplementary Fig. S3b); below-ground stems (rhizomes) that probably belonged to such plants yielded three statistically indistinguishable ages pooled as AD 1280–1390.

The tsunami deposit from event B probably exceeds the one from

1960 in thickness and landward extent. Because the 1837 tsunami was large in Hawaii^{12,16}, we expected this penultimate sand sheet to date from the early nineteenth century. Instead, a burned horizon mostly 2 cm below the sand dates to AD 1450–1510 or 1590–1620, as judged from four statistically equivalent ages on charred twigs. Because it followed the fire, probably by a century at most, we correlate event B with the extensive subsidence and devastating tsunami of 1575 (Fig. 1b).

We checked additional estuarine records in a further, futile search for signs of the 1837 earthquake. These records include trees that the 1960 earthquake lowered into tidal freshwater farther up the Río Maullín (red triangle, Fig. 1c). Residents on hand for the 1960 earthquake testify that a forest, green and emergent before the earthquake, lost its foliage from routine tidal submergence in the first few years thereafter. Several decades later, defoliated trunks dominated an area of 10 km². But several decades after the 1837 earthquake, a nautical chart¹⁹ depicted all trees in this area as leafy (Fig. 3a). In an accompanying report²⁰, the expedition botanist does not mention dead or dying trees among the forest's riparian plants and animals, which he studied for four days. We cut slabs of 15 dead standing trees in 2003 to estimate their lifespans by counting annual rings. We assume these trees died in 1960. In that case, ten of them were alive in 1837 and two in 1737 (Fig. 3b). This finding suggests that the forest failed to subside in 1837 as much as it did in 1960, in agreement with the nautical survey and the botanist's report.

Shoreline changes provide additional evidence that the 1837 earthquake did not produce 1960-size subsidence along the Río Maullín. Some of the islands and pastures that subsided in 1960 into the middle or lower part of the intertidal zone are barren intertidal or subtidal flats (Fig. 1c, green triangles). At a similar time after the 1837 earthquake, these areas were charted¹⁹ as emergent and vegetated (Supplementary Fig. S1b).

Earthquakes evident in these various estuarine records thus occurred less often than did earthquakes in the historical sequence: 1575, 1737, 1837, 1960 (Fig. 2c). The best-defined of the earthquake intervals recorded geologically, which together span most of the past millenium, average nearly 300 yr—more than double the historical average of 128 years. The 1960 earthquake ended a 385-year interval that includes the years 1737 and 1837. The poorly understood earthquakes of 1737 and 1837 probably released too little seismic moment midway along the 1960 rupture to leave tsunami deposits or subsidence stratigraphy at the Río Maullín.

Where size varies markedly among successive earthquakes on the same part of a fault, much of the fault slip during the largest earthquakes may have thus accumulated before earlier earthquakes of smaller size. Such storage through multiple recurrence intervals probably helps to explain the enormity of the 2004 Sumatra–Andaman earthquake. The fault slip in 2004 near the Nicobar Islands amounted to 10 m (ref. 21) in an area where the fault had last ruptured in 1881 during an earthquake of estimated magnitude 7.9 (ref. 22). By contrast, the fault loading between 1881 and 2004 amounted to less than 4 m at plate-convergence rates recently estimated from satellite geodesy²² and less than 7 m at rates inferred from long-term plate motions³. As in the 1960 Chilean case, the 2004 earthquake may thus have used accumulated plate motion that a previous earthquake left unspent.

Received 9 March; accepted 15 June 2005.

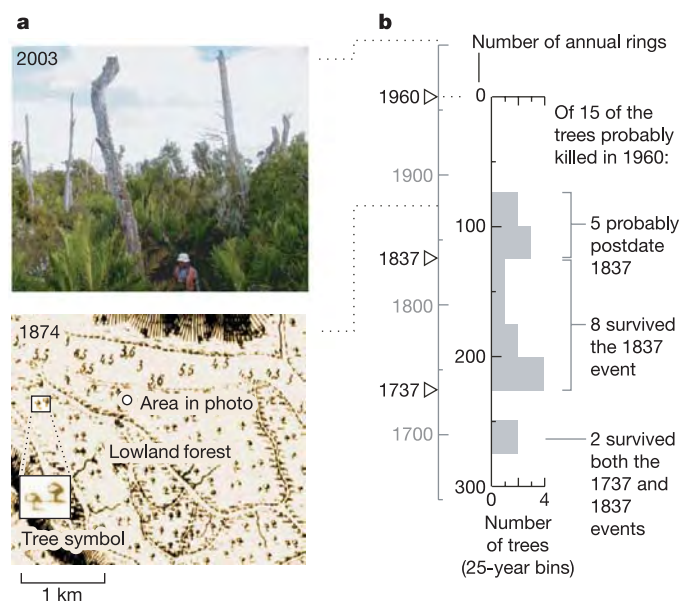


Figure 3 | Arboreal evidence for difference between the 1960 and 1837 earthquakes, in area shown by red triangle in Fig. 1c. (See Supplementary Fig. S1d, e). **a**, Views of riparian forest several decades after each earthquake (1874 image from ref. 19). **b**, Counts of annual rings in trees probably killed in 1960 (species, Supplementary Table S4).

- Shimazaki, K. & Nakata, T. Time-predictable recurrence model for large earthquakes. *Geophys. Res. Lett.* **7**, 279–282 (1980).
- Weldon, R., Scharer, K., Fumal, T. & Biasi, G. Wrightwood and the earthquake cycle: what a long recurrence record tells us about how faults work. *GSA Today* **14**(9), 4–10 (2004).
- DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. Current plate motions. *Geophys. J. Int.* **101**, 425–478 (1990).
- Lomnitz, C. Major earthquakes and tsunamis in Chile during the period 1535 to 1955. *Geol. Rundsch.* **59**, 938–960 (1970).
- Urrutia, R. & Lanza, C. *Catástrofes en Chile 1541–1992* 90 (Editorial La Noria, Santiago, Chile, 1993).

6. Plafker, G. & Savage, J. C. Mechanism of the Chilean earthquakes of May 21 and 22, 1960. *Geol. Soc. Am. Bull.* **81**, 1001–1030 (1970).
7. Plafker, G. Alaskan earthquake of 1964 and Chilean earthquake of 1960: implications for arc tectonics. *J. Geophys. Res.* **77**, 901–925 (1972).
8. Kanamori, H. & Cipar, J. J. Focal process of the great Chilean earthquake May 22, 1960. *Phys. Earth Planet. Int.* **9**, 128–136 (1974).
9. Cifuentes, I. The 1960 Chilean earthquakes. *J. Geophys. Res.* **94**, 665–680 (1989).
10. Barrientos, S. E. & Ward, S. N. The 1960 Chile earthquake: inversion for slip distribution from surface deformation. *Geophys. J. Int.* **103**, 589–598 (1990).
11. *El Maremoto del 22 de Mayo de 1960 en las Costas de Chile* (Servicio Hidrográfico y Oceanográfico de la Armada de Chile, Valparaíso, Chile, 1961).
12. Lander, J.F. & Lockridge, P.A. United States tsunamis 1690–1988. (National Geophysical Data Center Publ. 41–2, Boulder, Colorado, 1989).
13. Watanabe, H. *Comprehensive List of Destructive Tsunamis to Hit the Japanese Islands* [in Japanese] (Univ. Tokyo Press, Tokyo, 1998).
14. Lockridge, P. A. *Tsunamis in Peru-Chile* (Report SE-39, World Data Center A for Solid Earth Geophysics, Boulder, Colorado, 1985).
15. Ninomiya, S. Tsunami in Tohoku coast induced by earthquake in Chile; a chronological review. *Tohoku Kenkyu* [in Japanese with English summary] **10**, 19–23 (1960).
16. Abe, K. Size of great earthquakes of 1837–1974 inferred from tsunami data. *J. Geophys. Res.* **84**, 1561–1568 (1979).
17. Atwater, B. F., Jiménez, N. H. & Vita-Finzi, C. Net late Holocene emergence despite earthquake-induced submergence, south-central Chile. *Quat. Int.* **15/16**, 77–85 (1992).
18. Atwater, B. F. *et al.* Surviving a tsunami—lessons from Chile, Hawaii, and Japan. *US Geol. Surv. Circ.*, 1187 (1999).
19. Vidal Gormaz, F. Exploraciones hidrográficas practicadas en las costas de Chile, por la marina militar de la República (nautical chart). *Anuario Hidrográfico de la Marina de Chile, Año 1* (Imprenta Nacional, Santiago, Chile, 1875).
20. Juliet, C. Informe del ayudante de la comisión exploradora de Chiloé i Llanquihue. *Anuario Hidrográfico de la Marina de Chile, Año 1* 263–339 (Imprenta Nacional, Santiago, Chile, 1875).
21. Lay, T. *et al.* The great Sumatra-Andaman earthquake of 26 December 2004. *Science* **308**, 1127–1133 (2005).
22. Ortiz, M. & Bilham, R. Source area rupture parameters of the 31 December 1881 Mw = 7.9 Car Nicobar earthquake estimated from tsunamis recorded in Bay of Bengal. *J. Geophys. Res.* **108**(B4), 2215, doi:10.1029/2002JB001941 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by Chile's Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt) and by the US Geological Survey. Logistical help came from the Municipality of Maullín and its people (M. I. Silva, P. Soto, R. Vergara, J. Gallardo, G. Andrade, J. Soarzo and C. Ruiz), and from the Servicio Hidrológico y Oceanográfico of the Armada de Chile. The manuscript incorporates suggestions from S. Barrientos, S. Bondevik, C. Lomnitz, A. Nelson, K. Wang, J. Clague and E. Geist.

Author Contributions M.C. and B.A. led the fieldwork and writing. F.T. studied documents; Y.S. studied diatoms; G.M. studied tree slabs. M.L. and I.S. contributed to three seasons of fieldwork, G.M. and C.Y. to two, and A.E., M.H., T.K., J.K.M., C.P.R., Y.R. and M.S. to one.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.C. (marco.cisternas@ucv.cl).

LETTERS

Mate fidelity and intra-lineage polygyny in greater horseshoe bats

Stephen J. Rossiter^{1,2}, Roger D. Ransome², Christopher G. Faulkes¹, Steven C. Le Comber¹ & Gareth Jones²

Mating strategies that lead to increased kinship within socially cooperative groups may offer inclusive fitness benefits to individuals^{1–3}, but can also result in higher levels of inbreeding^{4–6}. Here we show in a sexually segregated bat species that females avoid this conflict through two mating behaviours. First, most females revisit and breed with specific, individual males across years, so that their single offspring born in different years are full siblings. Second, relatives in the maternal line, including mothers and daughters, share breeding partners (intra-lineage polygyny) more often than expected by chance. Although these behaviours increased levels of co-ancestry among colony members, there was no concomitant rise in inbreeding. We suggest that when females engage in mate fidelity and intra-lineage polygyny, kin ties among female roost mates will be strengthened, thereby potentially contributing to social group cohesiveness. Our findings reveal the hidden complexity that can underlie polygynous breeding, and highlight a new potential route by which female mate choice could influence social evolution.

The importance of kinship for the evolution and maintenance of social organization has been demonstrated in a wide range of taxa^{2,3,6}. Elevated genealogical relatedness has been linked to philopatry (the tendency to return to or remain near a particular site)^{4,7}, delayed dispersal⁸ and reproductive skew⁶. In contrast, few studies have considered the potential effect of female choice on kin structure, despite a wealth of evidence showing that females choose mates to enhance the fitness of their progeny⁹. Here we address this question by examining long-term breeding patterns in the sexually segregated greater horseshoe bat (*Rhinolophus ferrumequinum*). In many mammals, mate bonds are largely determined by the extent to which males can monopolize mates or resources¹⁰; however, studies of bats indicate a larger role for female choice¹¹. In greater horseshoe bats, mating occurs in both autumn (mid-August to November) and spring (April), at which times females disperse and form temporary multi-individual groups with single territorial males in caves¹². Sperm are stored through hibernation until fertilization occurs in April, after which females form single-sex summer colonies to raise their single pups¹². Females typically show life-long philopatry to their natal colony, promoting the aggregation of matrilineal relatives, although occasional immigrants may establish new matriline¹³. Male offspring disperse (~10 km, range 2–35 km) in their second year and remain segregated from females except when mating¹².

We genotyped microsatellite loci in all bats captured in our study population between 1991 and 2002 ($n = 452$), including ten cohorts of offspring (1993–2002), their mothers and a large proportion of candidate fathers sampled from all known roosts in the study area. We used maternity analysis together with 17 years' worth of data on female–offspring attachment, to identify mothers of 371 individual bats. Bats were allocated to matriline on the basis of oldest known matrilineal relatives, and mother–offspring pairs were used to

reconstruct matrilineal pedigrees of up to seven tiers spanning 16 years (1987–2002).

As female greater horseshoe bats produce one offspring annually, each bat represents the outcome of a separate mating. We identified breeding pairs using a likelihood method of paternity analysis¹⁴ to all bats of known maternity born since 1993 ($n = 348$), and identified fathers of 232 offspring (66.7%). The most likely father was assigned paternity at $\geq 80\%$ confidence in 204 cases (88% of offspring), out of which paternity was assigned at $\geq 95\%$ confidence in 140 cases (60%). Parentage of the 232 offspring assigned both parents was shared by a total of 75 females and 58 males. Individual females bred with identified males over an average ($\pm$ s.e.m., used here and throughout) of 3.09 ± 0.25 years (range 1–9), representing 2.40 ± 0.17 (range 1–7) partners per female. Male partners were then mapped onto pedigrees of females to examine patterns of pairings within and across years.

Mate fidelity was recorded in 56.8% (29/51) of females that produced pups in more than one year, 63.0% (24/38) of females producing pups in more than two years, and 77.7% (21/27) of females that bred in more than three years. In total, 34 parental combinations, made up of 27 females and 14 males, were repeated on up to five occasions (mean 2.6 ± 0.17 , range 2–5), spanning 2–8 years (Fig. 1). Randomizations showed that repeated pairings occurred more than expected by chance ($Z = -9.690$, $P < 0.0001$), and were more likely to be repeated in successive years ($Z = -2.676$, $P = 0.004$). Consequently, full siblings tended to be of a similar age and, therefore, full sisters could theoretically roost together for >20 years. As greater horseshoe bats show sexual segregation before mating, mate fidelity must involve the reunion of pairs, as reported for grey seal¹⁵ and some migratory birds¹⁶, rather than persistent mate bonds.

As males can occupy the same territory for several years¹², mate fidelity could arise from females' preferences for either individuals or sites. In 29 cases of mate fidelity, 23 (79.3%) involved site fidelity, one (3.4%) occurred after the male switched site, and in five instances (17.2%), neither scenario could be confirmed. We identified six examples of a female continuing to breed with the holder of a territory even after the male changed. However, in 15 cases when a female switched partner after a run of repeated matings, the original male was nearly always known to still be alive ($n = 11$) and often remaining in the same territory ($n = 6$). Therefore, mate fidelity does not seem to occur solely as an artefact of territory fidelity. However, whether female choice is based on mate or site fidelity, the extended association between males and territories means that in this species the ultimate consequences for social evolution will be the same.

Female matrilineal relatives mated with the same male (intra-lineage polygyny) more frequently than when mating was random ($Z = 1.744$, $P = 0.04$). We detected 20 groups of related females sharing mates, with 2–5 individuals per group (mean 3 ± 0.27),

¹School of Biological Sciences, Queen Mary, University of London, London E1 4NS, UK. ²School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK.

representing eight matriline. Examination of pairwise genealogical relationships among females engaging in intra-lineage polygyny revealed both lineal (descendent) ($n = 18$) and collateral (non-descendent) ($n = 25$) kin pairs, consisting of mothers and daughters ($n = 11$), grandmothers and granddaughters ($n = 7$), half-sisters ($n = 6$), and on the basis of matrilineal ancestry, half-aunts and half-nieces ($n = 10$), half-great-aunts and half-great-nieces ($n = 3$), half-great-great-aunts and half-great-great-nieces ($n = 1$), and half-first-cousins ($n = 5$) and half-first-cousins-twice-removed ($n = 1$). Although the adaptive origin of intra-lineage polygyny in greater horseshoe bats might lie in mate-choice copying or information transfer about mating sites, this behaviour leads to considerably higher relatedness among some colony members than expected from either typical polygynous or monogamous systems; intra-lineage polygyny among first-order and second-order relatives resulted in relatedness coefficients between their progeny increasing from 0.125 to 0.375 and from 0.0625 to 0.3125, respectively (Fig. 2).

Social unit models of gene dynamics show that intra-lineage polygyny can theoretically raise co-ancestry among discrete generations of females without inbreeding^{17–19}. When matrilineal relatives dominate breeding groups, genetic subdivision of lineages is expected with excess heterozygosity within lineages^{17,18}. To date, the effects of social organization on fixation indices have been examined predominantly in species with year-round male defence²⁰, in which intra-lineage polygyny could reflect a lack of spatial independence between living and mating, rather than active female choice. To test these predictions in a species with overlapping generations and sexual segregation, we calculated F -statistics and found significant differentiation among groups of matrilineal breeding females ($F_{st} = 0.143$, 95% confidence interval (CI) 0.112 to 0.170, $P < 0.001$) as well as higher levels of within-group heterozygosity than expected under Hardy–Weinberg equilibrium ($F_{is} = -0.163$, 95% CI -0.209 to -0.119 , $P < 0.001$). Genetic structure ($F_{st} = 0.081$, 95% CI 0.066 to 0.095, $P < 0.001$) and excess heterozygosity within groups ($F_{is} = -0.081$, 95% CI -0.103 to -0.058 , $P < 0.001$) were also detected for the females' offspring.

Several aspects of the described mating system might be expected to increase the potential for inbreeding, which is determined largely by the dispersion of male and female kin, and has thus been

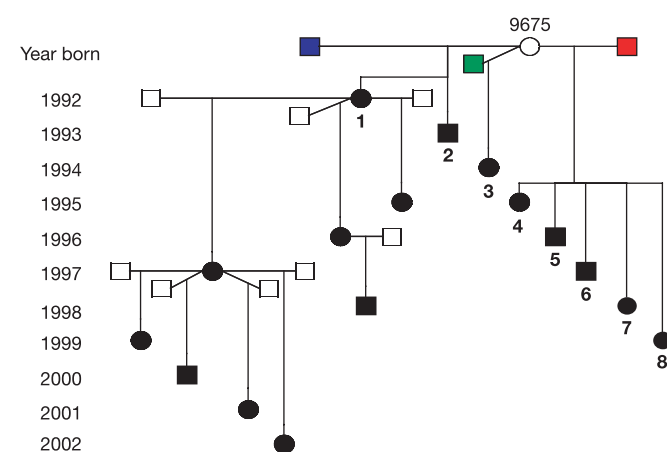


Figure 1 | Pedigree illustrating mate fidelity by female 9675. Breeding with three different males over eight years resulted in two full sibships, comprising two offspring (1 and 2) and five offspring (4–8). Circle, female; square, male. Individual males are coded by colour (blue, green and red). Black symbols represent offspring and open squares represent males that either sired one offspring in the pedigree ($n = 7$) or were unidentified ($n = 1$).

considered both a risk and an adaptive function of philopatry^{4–6}. In our study population, males often settle within the range over which females disperse for mating and sire a high proportion of offspring born in their native colony²¹. However, the coincidence of mate fidelity and intra-lineage polygyny might further increase the potential for inbreeding among close relatives, regardless of the males' origins. In theoretical models of intra-lineage polygyny, inbreeding is avoided by random male dispersal and replacement of adult females by their progeny each generation¹⁷. These assumptions do not apply to greater horseshoe bats, which, like many mammals, live in societies characterized by overlapping generations, with male reproductive spans typically exceeding one season. Consequently, shared mate choice among female relatives could promote inbreeding, particularly when coupled with mate fidelity.

To quantify the extent of consanguineous mating, we calculated individual inbreeding coefficients (f) from pedigrees. We identified 16 inbreeding events with generally low inbreeding coefficients (mean $f = 0.09 \pm 0.02$, range 0.016–0.25). Only one case of close inbreeding (0.38%; father–daughter) and six cases of moderate inbreeding (0.84%; three maternal half-siblings, two paternal half-siblings, one granddaughter–maternal grandfather) were detected. We found no difference in the observed number of pairings between bats sharing either a matrilineal ancestor or any ancestor and the distribution of values in the randomizations ($Z = 1.402$ for matrilineal ancestors, not significant (NS); $Z = 1.448$ for any ancestor, NS). Nor did we detect any increase in inbreeding when we compared the distribution of relatedness coefficients of breeding pairs in our data set to the null distribution generated under the assumption of random mating with respect to kinship, either when the analysis was restricted to males that bred in each year ($D = 0.059$, $n = 232$, NS), or when all males known to be alive were included ($D = 0.082$, $n = 232$, NS). Our results therefore show that mate fidelity and intra-lineage polygyny in this species cannot be explained by inbreeding avoidance, and are not associated with a concomitant rise in inbreeding.

Kinship is an important factor in the maintenance of cooperative behaviours^{1–3,22,23} and has been linked to sociality in a range of

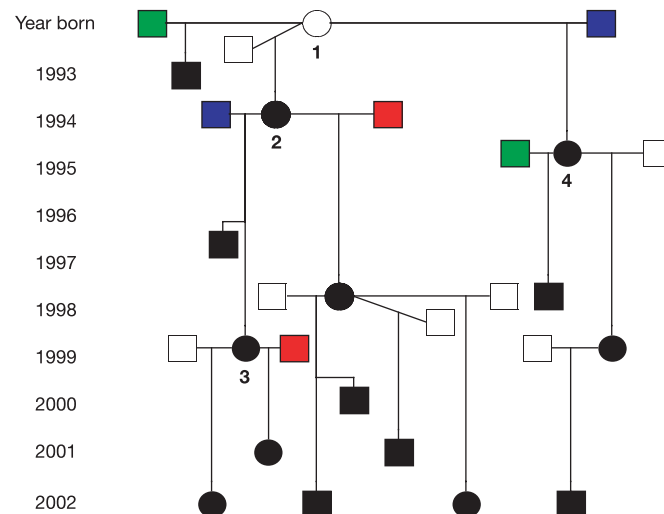


Figure 2 | Pedigree illustrating intra-lineage polygyny in matriline 8386. Circle, female; square, male. Individual males are coded by colour (blue, green and red). Black symbols represent offspring and open symbols represent males that either sired one offspring in the pedigree ($n = 6$) or were unidentified ($n = 1$). Female 1 shares mating partners (blue and green) with daughters 2 and 4, but in neither case does the daughter mate with her own father. Offspring 3 is the maternal half-niece of offspring 4, but these bats also share a father (blue) and therefore have a relatedness coefficient of 0.375.

mammals^{2,3,22,23}, including greater horseshoe bats¹³. In such cases, behaviours that lead to increased levels of relatedness within social groups while avoiding the potential costs of inbreeding are likely to become favoured by selection, ultimately accelerating the processes of population divergence and speciation that are associated with patterns of mammalian social organization^{20,24}.

METHODS

Study background and tissue collection. We studied a summer colony of about 34 (mean over ten years) breeding female greater horseshoe bats at Woodchester Mansion in Gloucestershire, UK (51°2' N, 2°90' W)^{13,25}. Since 1982, all adult females and offspring captured at this roost have been ringed. Adults of both sexes have also been ringed during annual surveys of underground sites (radius ≤ 25 km from the roost), including 30 mating territories. A territory was defined either as a small single cave or a discrete part of a larger underground system that was occupied by a single adult male roosting with one or more adult females during the breeding season¹². Between 1993 and 2002, we sampled wing membrane biopsies from ten complete cohorts of bats ($n = 348$) and 104 adults (51 males, 53 females). On the basis of recapture rates, we estimate that $>90\%$ of bats in this population have been ringed. Tissue punches were stored at -20°C in 90% ethanol or in 20% dimethylsulphoxide in saturated NaCl ($\sim 6\text{M}$).

DNA isolation and microsatellite genotyping. We isolated genomic DNA by either salt-chloroform precipitation or using the Qiagen DNeasy kit. All 452 sampled bats were genotyped at 15–19 (mean $\pm$ s.d. 17 ± 1.1) polymorphic microsatellite loci (GenBank accession numbers AF160200, AF160202, AF160205, AF160207, AF160208, AF160210, AF160211, AJ560695–AJ560698, AJ560702–AJ560704, AJ560708, AJ560710–AJ560713) (refs 26, 27). Mean heterozygosity was 0.60, and the mean number of alleles per locus was 7.11. Heterozygosity levels for these loci were previously shown not to deviate from Hardy–Weinberg equilibrium, indicating that null alleles are rare^{27,28}. Primers were 5'-fluoro-labelled, and 15 μl samples for polymerase chain reactions contained AmpliTaq Gold Taq polymerase (PE Applied Biosystems) and final primer concentrations of 0.667 μM . Amplified products were visualized on an ABI 3700 sequencer (PE Applied Biosystems) and allele sizes were analysed using the software GENESCAN v.3.1 and GENOTYPER v.3.6.

Maternity analysis and pedigree construction. We used the likelihood-based parentage analysis software CERVUS v.2.0 (ref. 14) to confirm the maternity of 268 offspring (born 1993–2002) assigned on the basis of attachment, and to match 80 unattached offspring to candidate lactating females. We additionally assigned mothers to 23 bats born before 1993 on the basis of postnatal attachment to a lactating female. Of these cases, maternity was confirmed by genetic analysis in all instances where both putative mother and offspring were genotyped as adults. Maternal sibships (siblings with the same mother) comprised 1–12 bats. We used data on all mother–offspring pairs to reconstruct pedigrees for all matrilineal lines using the program CYRILLIC (CyrillicSoftware).

Paternity analysis. For paternity analysis of 348 offspring (using CERVUS), we assumed that all males aged at least 2 yr were potential fathers, including those born in the colony. Females were assumed not to breed with their adult sons on the basis that they were never observed to share mating sites. We therefore removed sons from their respective mother's list of potential partners. No other male relatives were excluded, permitting potential father–daughter mating. To derive confidence estimates, we ran simulations assuming an average sampling rate of 75% on the basis of recapture rates within the study area, and a typing error rate of 1%. For each offspring, paternity was assigned to the male with the highest likelihood ratio, providing that the male mismatched at no more than one locus (or, for cases in which a mismatch occurred at the same locus in the mother–offspring pair, at no more than two loci). No differences were observed in the numbers of paternities awarded to Woodchester-born and immigrant males (S.J.R., R.D.R., C.G.F., Dawson, D.A. & G.J., submitted manuscript). Mean log-likelihood ratios (LOD scores) and Delta values were 7.27 and 3.92, respectively, and total exclusionary power for paternity assignment on the basis of 19 markers was 0.999.

Statistical analysis. Randomizations were used to assess the significance of observed levels of mate fidelity, serial mate fidelity, intra-lineage polygyny and inbreeding (see Supplementary Information).

Individual coefficients of inbreeding (f) were calculated from pedigree relationships among 420 bats using the software PEDIGREE VIEWER. We estimated the percentage frequency of different categories of close ($f \geq 0.25$) and moderate ($0.25 \geq f \geq 0.125$) inbreeding events based on the number of each category that could have been detected from the pedigree²⁹.

On the basis of ten pooled cohorts, the mean number of alleles per locus

(allelic diversity) was 6.94. We calculated Weir–Cockerham F -statistics and assessed their significance using the software GENETIX. Multi-locus F_{IS} estimates derived separately for ten cohorts showed no significant direction or deviation from Hardy–Weinberg expectations (-0.031 to 0.036 , $P > 0.05$). We examined genetic structure among groups of matrilineal females containing 5 ± 0.62 individuals (range 3–11, $n = 12$ groups), and among their groups of progeny (12.83 ± 2.41 individuals, range 4–43, $n = 18$ groups).

Received 16 May; accepted 24 June 2005.

- Hamilton, W. D. The genetical evolution of social behaviour. *J. Theor. Biol.* **7**, 1–52 (1964).
- Hepper, P. G. in *Kin Recognition* (ed. Hepper, P. G.) 1–5 (Cambridge Univ. Press, Cambridge, 1991).
- Chapais, B. & Berman, C. M. *Kinship and Behavior in Primates* (Oxford Univ. Press, New York, 2004).
- Shields, W. M. *Philopatry, Inbreeding, and the Evolution of Sex* (State Univ. New York Press, Albany, 1982).
- Pusey, A. E. Sex-biased dispersal and inbreeding avoidance in birds and mammals. *Trends Ecol. Evol.* **2**, 295–299 (1987).
- Reeve, H. K., Westneat, D. F., Noon, W. A., Sherman, P. W. & Aquadro, C. F. DNA fingerprinting reveals high levels of inbreeding in colonies of the eusocial naked mole-rat. *Proc. Natl Acad. Sci. USA* **87**, 2496–2500 (1990).
- Greenwood, P. J. Mating systems, philopatry and dispersal in birds and mammals. *Anim. Behav.* **28**, 1140–1162 (1980).
- Emlen, S. T. An evolutionary theory of the family. *Proc. Natl Acad. Sci. USA* **92**, 8092–8099 (1995).
- Andersson, M. *Sexual Selection* (Princeton Univ. Press, Princeton, 1994).
- Clutton-Brock, T. H. Mammalian mating systems. *Proc. R. Soc. Lond. B* **236**, 339–372 (1989).
- McCracken, G. F. & Wilkinson, G. S. in *Reproductive Biology of Bats* (eds Crichton, E. G. & Krutzsch, P. H.) 321–362 (Academic, New York, 2000).
- Ransome, R. D. in *The Handbook of British Mammals* 3rd edn (eds Corbet, G. B. & Harris, S.) 88–94 (Blackwell, Oxford, 1991).
- Rossiter, S. J., Jones, G., Ransome, R. D. & Barratt, E. M. Relatedness, kin-biased foraging in the greater horseshoe bat *Rhinolophus ferrumequinum*. *Behav. Ecol. Sociobiol.* **51**, 510–518 (2002).
- Marshall, T. C., Slate, J., Kruuk, L. & Pemberton, J. M. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* **7**, 639–655 (1998).
- Amos, B., Twiss, S., Pomeroy, P. & Anderson, S. Evidence for mate fidelity in the gray seal. *Science* **268**, 1897–1899 (1995).
- Sandercock, B. K., Lank, D. B., Lanctot, R. B., Kempenaers, B. & Cooke, F. Ecological correlates of mate fidelity in two Arctic-breeding sandpipers. *Can. J. Zool.* **78**, 1948–1958 (2000).
- Chesser, R. K. Gene diversity and female philopatry. *Genetics* **127**, 437–447 (1991).
- Chesser, R. K. Influence of gene flow and breeding tactics on gene diversity within populations. *Genetics* **129**, 573–583 (1991).
- Sugg, D. W., Chesser, R. K., Dobson, F. S. & Hoogland, J. L. Population genetics meets behavioural ecology. *Trends Ecol. Evol.* **11**, 338–342 (1996).
- Storz, J. F. Genetic consequences of mammalian social structure. *J. Mammal.* **80**, 553–569 (1999).
- Rossiter, S. J., Jones, G., Ransome, R. D. & Barratt, E. M. Parentage, reproductive success and breeding behaviour in the greater horseshoe bat (*Rhinolophus ferrumequinum*). *Proc. R. Soc. Lond. B* **267**, 545–551 (2000).
- Griffin, A. S. & West, S. A. Kin discrimination and the benefit of helping in cooperatively breeding vertebrates. *Science* **302**, 634–636 (2003).
- Wilkinson, G. S. Reciprocal food sharing in the vampire bat. *Nature* **308**, 181–184 (1984).
- Bush, G. L., Case, S. M., Wilson, A. C. & Patton, J. L. Rapid speciation and chromosomal evolution in mammals. *Proc. Natl Acad. Sci. USA* **74**, 3942–3946 (1977).
- Ransome, R. D. Earlier breeding shortens life in female greater horseshoe bats. *Phil. Trans. R. Soc. Lond. B* **350**, 153–161 (1995).
- Rossiter, S. J., Burland, T. M., Jones, G. & Barratt, E. M. Characterization of microsatellite loci in the greater horseshoe bat *Rhinolophus ferrumequinum*. *Mol. Ecol.* **8**, 1957–1969 (1999).
- Dawson, D. A., Rossiter, S. J., Jones, G. & Faulkes, C. G. Microsatellite loci for the greater horseshoe bat, *Rhinolophus ferrumequinum* (Rhinolophidae, Chiroptera) and their cross-utility in 17 other bat species. *Mol. Ecol. Notes* **4**, 96–100 (2004).
- Rossiter, S. J., Jones, G., Ransome, R. D. & Barratt, E. M. Genetic variation and population structure in the endangered greater horseshoe bat (*Rhinolophus ferrumequinum*). *Mol. Ecol.* **9**, 1131–1135 (2000).
- Marshall, T. C. *et al.* Estimating the prevalence of inbreeding from incomplete pedigrees. *Proc. R. Soc. Lond. B* **266**, 1533–1539 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Woodchester Mansion Trust for access to

the roost and the many volunteers who helped with data collection. We thank T. Burland, D. Dawson, B. Manley, C. Mein and R. Nichols for advice and technical support, and A. Bourke, A. Hildrew, R. Nichols, A. Overall, D. Polly and F. Ratnieks for helpful comments on an earlier version of the manuscript. Bats were caught and sampled under license from English Nature and the Home Office. This work was funded by the Natural Environment Research Council (NERC), and microsatellite development was supported by the NERC Sheffield Molecular Genetics Facility.

Author Contributions S.J.R. conceived the project with G.J., undertook the

laboratory work and wrote the paper. R.D.R. undertook most of the fieldwork with help from G.J. S.J.R. and C.G.F. jointly analysed the pedigree and genetic data, and S.L.C. wrote the computer programs to implement the randomizations and statistical tests.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.J.R. (s.j.rossiter@qmul.ac.uk).

LETTERS

A sensory source for motor variation

Leslie C. Osborne¹, Stephen G. Lisberger^{1,2} & William Bialek³

Suppose that the variability in our movements^{1–9} is caused not by noise in the motor system itself, nor by fluctuations in our intentions or plans, but rather by errors in our sensory estimates of the external parameters that define the appropriate action. For tasks in which precision is at a premium, performance would be optimal if no noise were added in movement planning and execution: motor output would be as accurate as possible given the quality of sensory inputs. Here we use visually guided smooth-pursuit eye movements in primates¹⁰ as a testing ground for this notion of optimality. In response to repeated presentations of identical target motions, nearly 92% of the variance in eye trajectory can be accounted for as a consequence of errors in sensory estimates of the speed, direction and timing of target motion, plus a small background noise that is observed both during eye movements and during fixations. The magnitudes of the inferred sensory errors agree with the observed thresholds for sensory discrimination by perceptual systems, suggesting that the very different neural processes of perception and action are limited by the same sources of noise.

Smooth-pursuit eye movement is the familiar ‘tracking’ behaviour elicited by the motion of small targets across the visual field (Fig. 1). It is convenient to describe pursuit eye movements as depending on the speed and direction of target motion. However, the brain has no independent knowledge of these parameters, and must estimate them visually if experiments are designed to remove opportunities for prediction. Furthermore, although tracking over long timescales involves feedback and is driven by a combination of retinal and extra-retinal signals, the eye trajectory in the ~125-ms time interval before feedback can arrive is generated purely from estimates of the target’s motion, using visual inputs present before the onset of the response¹¹. At least for perception, these estimates are not perfect: humans and non-human primates can make reliable visual discriminations only among trajectories that differ by ~10% in speed and ~2–3° in direction^{12–17}. Perceptual discrimination thresholds are limited by noise in neural activity in sensory areas: estimates of speed and direction will fluctuate from trial to trial as the brain tries to decode this noisy representation^{18,19}.

If the brain’s estimate of speed on one trial is larger than the mean (as the result of noise), then the goal of the movement on this trial will be indistinguishable from that for a genuinely faster target speed, and the corresponding commands to the eyes will drive proportionately larger eye accelerations and velocities. Similar considerations apply to errors in direction. Figure 1a shows an ensemble of eye-velocity trajectories generated by random scalings and rotations of the mean trajectory of eye velocity for target motion that steps from 0 to 20° s^{–1} in a rightward direction rotated 9° above the horizontal; the standard deviations of these scalings and rotations were chosen to match the sensory noise levels of 10% and 2.3°. For comparison, in Fig. 1b we show an ensemble of actual pursuit trajectories in response to repeated presentations of the same direction of target motion; these trials are interspersed with target motions in other directions to

eliminate the possibility of prediction and to force pursuit to be guided by estimates of visual motion. Comparison of the synthetic and actual trajectories reveals that noise on the scale that limits perceptual discrimination is sufficient to generate variation in motor output that is close to what we see experimentally.

To generate appropriate motor outputs, the brain must represent when the target starts moving, in addition to estimating the speed and direction of target motion. In our experiments, target motion begins at a random time relative to the onset of a fixation spot. On average, the trajectory of smooth pursuit seems to be locked to the trajectory of the target. On a given single trial, however, the brain lacks a perfect marker of the time of target motion onset, and so it must be estimated. In contrast to direction and speed, little is known about the limits of perceptual discrimination of motion timing. In Fig. 1c we illustrate the consequences of errors in timing estimation. A standard deviation of just 15 ms produces a variation in eye movement trajectories that is larger than we see in experiments (Fig. 1b), suggesting that the timing of target motion must be represented with a precision of better than 15 ms.

The results of Fig 1a–c motivate the hypothesis that variability in smooth-pursuit trajectories is dominated by errors in sensory estimation. More formally: imagine that there is an ideal (vector) eye-velocity trajectory $v_{\text{ideal}}(t; t_0, v, \theta)$ in response to a target that starts to move at time t_0 at speed v and in direction θ . On any single trial, the brain has access only to noisy estimates of these parameters so that it makes errors δt_0 , δv and $\delta \theta$. Then, the actual eye movements will be

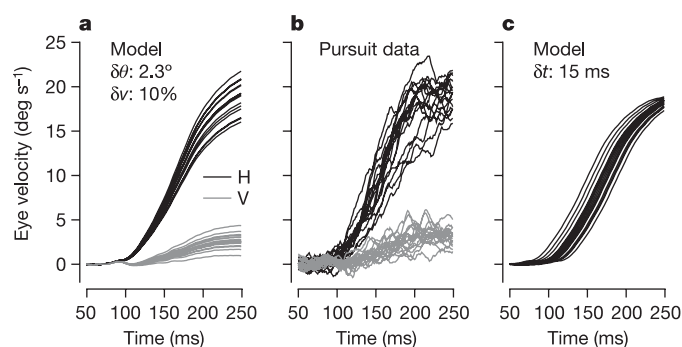


Figure 1 | Example of the variability in pursuit for a given target motion. **a**, Model data created from the mean pursuit-velocity time course, averaged over 184 repetitions of the same target motion. For each trace, the mean trajectory was rotated and scaled by a gaussian distributed ‘noise’ value, the standard deviation of which matches perceptual discrimination threshold values for direction and speed in human subjects (2.3° and 10%). **b**, Actual data showing 18 individual pursuit trials. **c**, Model data created by taking the same mean pursuit trajectory and jittering its start time by a gaussian distributed shift value with a standard deviation of 15 ms. Black and grey lines in **a** and **b** distinguish the horizontal (H) and vertical (V) components of eye velocity; only horizontal eye velocity is shown in **c**. Time is measured relative to target motion onset.

¹Sloan-Swartz Center for Theoretical Neurobiology, W. M. Keck Foundation Center for Integrative Neuroscience, Department of Physiology, and ²Howard Hughes Medical Institute, University of California at San Francisco, San Francisco, California 94143-0444, USA. ³Joseph Henry Laboratories of Physics and the Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA.

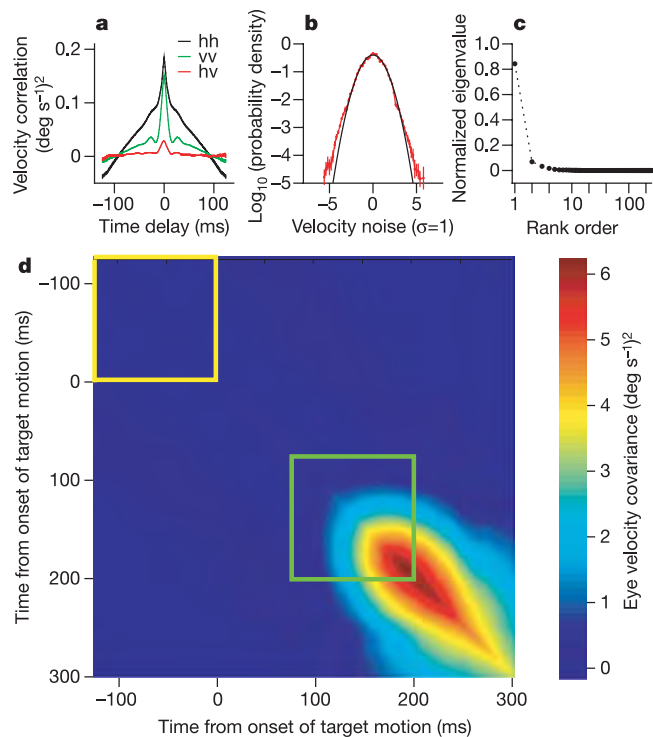


Figure 2 | Analysis of variation in pursuit trajectory for a single day's experiment. **a**, Temporal structure of correlation in eye-velocity variations before the onset of pursuit. Colours in key and traces labelled hh, vv and hv compare horizontal or vertical eye velocity to themselves or to each other (hv). **b**, Logarithm of probability density (red) and the best-fitting Gaussian curve (black) for the variations in eye velocity (in units of standard deviation, σ) before the onset of target motion. Error bars are s.d. divided by the mean. **c**, Rank order of the 250 normalized eigenvalues for ΔC . Standard deviations are smaller than the size of the symbols. **d**, Covariance matrix showing how the variation in horizontal eye velocity at any given time was related to that at all other times.

$\mathbf{v}_{\text{ideal}}(t; t_0 + \delta t, v + \delta v, \theta + \delta \theta)$. We assume that errors are small, so that we can approximate the consequences of changing parameters just by the first term in a Taylor series. Also, closer inspection of Fig. 1b reveals that the trial-to-trial fluctuations in trajectory include a more rapidly fluctuating component that is 'background noise', visible even before the initiation of pursuit: $\delta \mathbf{v}_{\text{back}}(t)$. Putting the terms together, we formalize the predicted trajectory in a natural, sensory space for a single trial as:

$$\begin{aligned} \mathbf{v}(t) = & \mathbf{v}_{\text{ideal}}(t; t_0, v, \theta) + \delta t_0 \cdot \frac{\partial \mathbf{v}_{\text{ideal}}(t; t_0, v, \theta)}{\partial t_0} \\ & + \delta \theta \cdot \frac{\partial \mathbf{v}_{\text{ideal}}(t; t_0, v, \theta)}{\partial \theta} + \delta v \cdot \frac{\partial \mathbf{v}_{\text{ideal}}(t; t_0, v, \theta)}{\partial v} \\ & + \delta \mathbf{v}_{\text{back}}(t) \end{aligned} \tag{1}$$

In equation (1), $\mathbf{v}_{\text{ideal}}$ can be recovered by averaging the actual trajectories over many trials, and the various derivatives can be extracted from the data without any further assumptions (see Methods for details).

The predictions of equation (1) can be tested by examining the covariance matrix of trial-to-trial fluctuations in eye velocity, shown in Fig. 2d. At times before the initiation of pursuit (yellow square), the covariance matrix should describe the background noise $\delta \mathbf{v}_{\text{back}}(t)$. After the initiation of pursuit (green square), there should be exactly three additional components that reflect the variances in δt_0 , $\delta \theta$ and δv . Experimentally (see Methods), we sampled the horizontal and vertical components of the vector velocity $\mathbf{v}(t)$ with 1-ms resolution throughout a 125-ms window after the initiation of pursuit, so that a single trajectory is described by 250 numbers. The covariance matrix of the background noise has the symmetric structure expected for stationary fluctuations, with ~ 80 eigenvalues that are significantly different from zero. In the time domain, the noise has significant components with a correlation time of less than 10 ms (Fig. 2a) and the distribution of noise velocities is nearly gaussian (Fig. 2b). In contrast, the matrix ΔC formed by subtracting the background covariance matrix (Fig. 2d, yellow box) from that for the first 125 ms of pursuit (Fig. 2d, green box) has just three eigenvalues that are significantly different from zero (Fig. 2c). The eigenvectors corresponding to these eigenvalues span the same three-dimensional space defined by the three derivatives of $\mathbf{v}_{\text{ideal}}(t)$ in equation (1). As summarized in Table 1, $93 \pm 1.4\%$ of the variance in trajectories is captured by these three eigenvectors, which in turn have 96–99% overlap with axes corresponding to errors in estimating target speed, direction and timing.

The observation of just three significantly non-zero eigenvalues for ΔC means that the variability of smooth-pursuit trajectories is effectively limited to three dimensions. As explained in the Methods, this collapse of dimensionality is enormously unlikely to have occurred by chance. Even though a number of motor behaviours have been shown to have similar low-dimensional structures^{5–7}, several aspects of our results seem novel. First, the low dimensionality cannot be interpreted as a limitation of the motor system itself, as the eye movement motor system is observed to generate trajectories that fill ~ 80 dimensions under the different conditions of fixation before the onset of pursuit. Second, the particular three dimensions in which the system operates are not arbitrary, but in fact are those predicted in advance. Finally, the magnitudes of the fluctuations in the three relevant directions have a clear physical and biological meaning in relation to the parameters used to specify visual motion.

In Fig. 3, we show how the three dimensions corresponding to speed, direction and timing errors can be used to synthesize the eye-movement trajectory on a single trial. Starting with the mean trajectory (Fig. 3a, dashed traces), we add components for each of the three natural modes (Fig. 3b), scaled by particular values of δv , $\delta \theta$ and δt_0 (arrowheads in Fig. 3c) to create accurate predictions (Fig. 3a, red and blue traces) of eye-velocity responses from individual trials (Fig. 3a, solid black and grey traces). We derived distributions of the values of δv , $\delta \theta$ and δt_0 by performing the same projection for each

Table 1 | Relationship between the eigenvectors of the covariance matrix ΔC and the axes of target direction, speed and motion-onset time

	Fraction of total variance	Fractional overlap with rotation $\mathbf{v}_{\text{dir}}(t)$ (direction noise)	Fractional overlap with scaling $\mathbf{v}_{\text{speed}}(t)$ (speed noise)	Fractional overlap with time shifts $\mathbf{v}_{\text{time}}(t)$ (timing noise)
Mode 1	0.7409 ± 0.0064	0.0001 ± 0.0002	0.9343 ± 0.005	0.8684 ± 0.0068
Mode 2	0.1150 ± 0.004	0.0062 ± 0.0057	0.0604 ± 0.005	0.1141 ± 0.0064
Mode 3	0.0586 ± 0.0026	0.9594 ± 0.0093	0.0013 ± 0.0008	0.0000 ± 0.0006
Total (3 modes)	0.9145 ± 0.0523	0.9657 ± 0.0062	0.9960 ± 0.0071	0.9825 ± 0.0093
Average ($n = 9$)	0.9326 ± 0.0143	0.9557 ± 0.0255	0.9961 ± 0.0097	0.9725 ± 0.0152

The top four rows show the analysis of the three eigenvectors or 'modes' that accounted for the largest percentage of the variance of pursuit, and the totals across the three largest modes for single experiment. The bottom row shows the average of the totals across the three largest modes for nine experiments. From left to right, columns give the percentage of the total variance, and the projections from the natural axes describing errors in the estimates of direction, speed and time onto the eigenvectors of the covariance matrix ΔC . Errors indicate standard deviations. Data set pk032404.

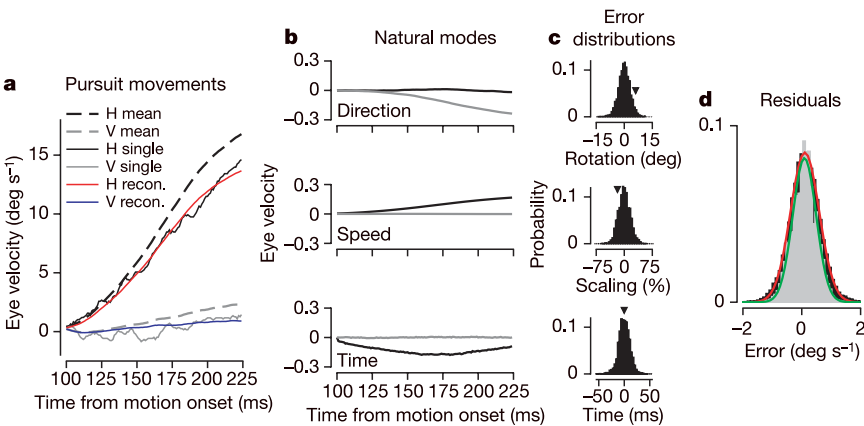


Figure 3 | Reconstruction of individual pursuit trials from the model described by equation (1). **a**, Eye velocity as a function of time for the mean trajectory, and for the actual and reconstructed trajectory for a single trial. **b**, Time courses of the sensory noise modes (v_{dir} , v_{speed} , v_{time}) in units of eye velocity per equivalent sensory error. **c**, Distributions of $\delta\theta$, δv and δt_0 for 184 responses to the same target trajectory. Arrowheads indicate the values of the errors used to reconstruct the single trial in **a**. **d**, Distributions of difference between actual and predicted eye velocity during pursuit (black) and the total noise present during fixation (grey), along with best-fitting gaussian functions (red, green).

individual pursuit response in the data set onto the natural modes (Fig. 3c); the distributions are approximated well by gaussians. Furthermore, the residual differences between actual and predicted responses have a distribution that agrees substantially with the distribution of background noise (grey versus black distributions in Fig. 3d). The agreement between the distributions of the background noise and the residuals of trial-by-trial reconstructions from the three natural modes is a restatement of our results on the eigenvalues and eigenvectors of ΔC , but presents the results in a different, and perhaps more intuitive, form.

In a total of nine experiments with three monkeys (pk, yo and wt) direction and speed errors had standard deviations of 2.1–3.5° and 11–18%, respectively. Each daily experiment involved a total of more than 1,000 trials, so statistical errors within a single experiment are much smaller than variations among experiments. The precision of pursuit behaviour correlates well with the results from perceptual experiments^{15,20}. For brief stimulus presentations like the ones used here (~125 ms), human subjects^{12–13} have thresholds of $\Delta\theta \approx 2.3^\circ$ in direction and $\Delta v/v \approx 10\%$ in speed. Longer stimulus presentations have been used for perceptual experiments in monkeys, yielding thresholds nearly identical to those in humans^{16–17}. Within the bounds of measurement error and differences in stimulus presentation, we conclude that the limits in precision for pursuit and perception are very similar.

For 200-ms stimulus presentations, human discrimination thresholds^{12,14–15} and the limits to precision inferred from the variability of pursuit trajectories improve, and are even more similar. However, on timescales longer than the ~125 ms we have considered,

the sensory–motor feedback loop for pursuit has been closed, complicating the comparison. For example, studies of perceptual and closed-loop motor variability in smooth pursuit reach opposite conclusions: both find that perception and pursuit have similar degrees of variability, but one addressing motion direction finds evidence for a common noise source²⁰, whereas another experiment addressing motion speed errors does not¹⁵. Because steady-state tracking is driven largely by extra-retinal signals, the absence (or presence) of covariation between perceptual and pursuit errors during steady-state tracking does not speak to the question of whether the errors in perceptual and motor readouts of visual motion arise from the same noise source.

The analysis of timing errors deserves special consideration. As hinted at by the results in Fig. 1c, the standard deviation of these timing errors is quite small, 7–10 ms across our set of nine experiments (Table 2). This is much smaller than the range of reaction times for discrete movements such as saccades²¹. Thus, even though the time of onset of target motion is not known to the brain *a priori*, the pursuit system is able to estimate that time with remarkable precision. Given the small number of spikes that are emitted by neurons in the MT region of the visual cortex in response to the first 100 ms of target motion²², the neural mechanisms that decide when to initiate a movement^{23,24} must be able to do so on the basis of the timing of just a few spikes.

We can think of the timing error δt_0 as a measure of latency relative to the mean that looks for the best fit of a template to the whole 125-ms open-loop segment of the eye-velocity trajectory rather than (for example) the traditional measure of latency as the moment at which

Table 2 | Limits to precision in the sensory estimates driving smooth pursuit

Experiment	Target speed (° s ⁻¹)	Fraction of variance	$\delta\theta_{\text{rms}}$ (°)	$\delta v_{\text{rms}}/v$	δt_{rms} (ms)	$C(\delta v; \delta t)$
pk121603	20	0.915 ± 0.002	2.48 ± 0.04	0.129 ± 0.002	8.3 ± 0.1	0.09 ± 0.02
pk032404	20	0.904 ± 0.005	2.34 ± 0.05	0.123 ± 0.002	7.9 ± 0.1	0.23 ± 0.01*
pk041004	20	0.924 ± 0.002	2.99 ± 0.05	0.158 ± 0.002	10.8 ± 0.2	0.25 ± 0.02*
wt112102	20	0.894 ± 0.004	2.86 ± 0.1	0.173 ± 0.005	10.2 ± 0.3	0.29 ± 0.03*
yo082603	15	0.932 ± 0.005	2.40 ± 0.07	0.177 ± 0.004	9.4 ± 0.3	0.10 ± 0.03
yo082203	10	0.936 ± 0.005	3.5 ± 0.1	0.181 ± 0.003	9.5 ± 0.3	0.07 ± 0.03
yo082703	25	0.929 ± 0.003	2.33 ± 0.06	0.172 ± 0.002	10.8 ± 0.2	0.22 ± 0.02*
pk021504	30	0.915 ± 0.002	2.14 ± 0.03	0.112 ± 0.002	8.3 ± 0.1	0.29 ± 0.02*
pk032504	10	0.913 ± 0.01	3.45 ± 0.08	0.162 ± 0.003	8.4 ± 0.2	− 0.08 ± 0.03

Each row shows data for one experiment analysing 125 ms of pursuit initiation. Columns show the target speed for the experiment; the fraction of the total variance accounted for by the noise attributed to sensory estimates of target direction, speed and the time of motion onset; the limits to precision in direction, speed and time defined by the root mean squared (rms) variation along the axis defined by each parameter; and the correlation between speed and timing noise (statistical significance indicated by asterisk). Target directions ranged within ±9° for all experiments except wt112102, which had a range of ±6°. Correlations were zero between direction and speed noise, and between direction and timing noise. Errors indicate standard deviations.

the eye velocity rises significantly above background noise. Although correlations between variations in direction and speed would imply a 'handedness' to the pursuit system that seems implausible, there is no symmetry that forbids correlations between variations in timing and speed. Indeed, we observe significant speed–timing correlations in many experiments (Table 2). The sign of the correlations corresponds to 'start later–go faster', but because we are analysing the open-loop response of pursuit, there is no feedback signal to ensure that late starts are compensated for by larger eye accelerations. Thus, it seems likely that the correlation between variations in speed and timing is intrinsic to the estimation or representation of motion in the visual system. In decomposing errors into direction, speed and timing components, it is important to account (as we do) for the fact that although direction and speed errors make orthogonal contributions to the pursuit trajectory, speed and timing errors do not. Instead, they point oppositely along similar axes in the three-dimensional space. This explains why speed and time have similar magnitude projections onto modes 1 and 2 of ΔC , but with a sign difference that disambiguates 'moving faster' from 'starting earlier'.

Other studies have shown that eye movements and perception share neural pathways and have access to the same sensory estimates of visual motion^{13,14,16–18,20,25}. Demonstrations of the limits to perceptual discrimination of target direction and speed have revealed that there is noise in the sensory inputs. We have shown that essentially the entire motor variation that is specific to pursuit lies along the axes of the sensory parameters of target direction, speed and timing, and that the limits to precision of pursuit are nearly the same as those for perception. An appealing and simple conclusion is that the initial response of the pursuit system adds little additional noise beyond the variations in sensory estimates, and thus its precision is defined by the noise in the sensory representations. Our findings do not indicate whether the precision of sensory representations of visual motion is limited by noise arising in the retina, or whether it accumulates along motion-sensitive neural pathways. Whatever its origin, we imagine that the variability we observe in pursuit initiation is reflected in the responses of cortical neurons (for example, in MT/V5).

We have predicted the structure of variations in the initiation of pursuit from first principles, and have provided data that are consistent with, but do not prove, the hypothesis that variation in initiation of pursuit arises largely from the sensory representation of visual motion. Further testing is needed to rule out alternatives in which a precise representation of target motion is degraded by noise that accumulates independently along perceptual and motor pathways. These alternatives require that noise added in the motor system preserve the sensory form and have a magnitude similar to that measured for perception. In the simplest concrete alternative, motor output variation is dominated by trial-by-trial fluctuation in the strength of commands sent to the eye muscles. But this model does not provide an explanation of the low-dimensionality of the noise, except by assuming a similar low dimensionality in the gain noise—that is, that the fluctuating components of the commands affect the entire 125-ms trajectory uniformly. More detailed computations show that the magnitude of the gain fluctuations must be tuned differently for each direction of motion to account quantitatively for the data. Finally, any model that ascribes the observed behavioural variability largely to the motor side of the nervous system must explain why the inferred gain variations are so large when the variability of motor neuron discharge is so small²⁶.

The overlap of the significant dimensions of pursuit variation with those expected from the parameters of the motion trajectory may have important implications for the operating principles of the brain's motor circuits. The agreement between the limits to precision in pursuit and perceptual behaviour biases us to think that sensory processing is the main contributor to variability in pursuit trajectory, and that other sources of noise in the system are effectively smaller,

perhaps because motor strategies are selected to minimize other noise sources^{4,8}. That variation in pursuit behaviour can be assigned largely to noise in representation of the sensory stimulus may fit with other examples in which the nervous system achieves optimal or near-optimal performance^{27–29}.

METHODS

Eye movements were recorded¹¹ from three male rhesus monkeys (*Macaca mulata*) that had been trained to fixate and track visual targets. Experiments lasted 2–3 h, during which the monkey sat in a specialized primate chair with its head immobilized, and received a juice or water reward for accurately tracking visual targets presented on a screen in front of it. All procedures had been approved by the Institutional Animal Care and Use Committee of the University of California, San Francisco and were in compliance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals.

The visual target was typically a 0.8° square spot presented in a dimly lit room on a high-resolution analogue display oscilloscope that subtended a 48° by 38° visual angle. Experiments were presented as a series of trials, each representing a single target motion. Each trial began with the monkey fixating a stationary target at centre-screen for a random interval of 700–1,200 ms. The target then underwent a step-ramp motion¹¹ with steps of 2.5–3.7° and ramped back towards the extinguished fixation point at a constant speed, typically 20° s^{−1}. Directions were chosen randomly from up to 14 directions (that is, −9° to +9° relative to horizontal, in 3° increments). Parameters of target motion were varied so that they were presented in random order.

Vertical and horizontal eye-velocity signals were passed through an analogue double-pole, low-pass filter that differentiated frequencies below 25 Hz and rejected higher frequencies with a roll-off of 20 dB per decade. Eye-position and velocity signals were sampled and stored at 1 kHz. Before analysis, each trial record was inspected and rejected if a saccade occurred within the time-window chosen for analysis. Data sets consisted of eye velocity responses to 112–223 repetitions of target motion in each direction. The time-window for analyses of 'background' data began 125 ms before target motion and ended with the onset of target motion. The 125-ms time-window for pursuit analyses began at eye-movement onset, determined by the intersection of two lines each fitted to pre- and post-pursuit intervals of average responses. Standard deviations were computed from analyses based on 40–50 random draws of half of the data set.

To recover the ideal trajectory $\mathbf{v}_{\text{ideal}}(t; t_0, \nu, \theta)$ (equation (1)), we averaged eye velocity over many responses to the same target motion. To compute the derivatives in equation (1), we took advantage of symmetries. First, changing the onset time t_0 should be equivalent to translating the response along the time axis of the ideal trajectory, so that $\mathbf{v}_{\text{time}} = \partial \mathbf{v}_{\text{ideal}} / \partial t_0 = -\partial \mathbf{v}_{\text{ideal}} / \partial t$. Second, changes in target speed should produce ideal trajectories that are uniformly scaled to be proportionately faster or slower, at least over a narrow dynamic range¹¹, so that $\mathbf{v}_{\text{speed}} = \partial / \partial \nu [(\nu / \nu_0) \mathbf{v}_{\text{ideal}}]$. Finally, changes in target motion direction should produce rotations of the ideal trajectory. We checked this last symmetry using principal component analysis of mean trajectories in response to (typically) 14 different directions. As expected if the changing target direction simply rotates the ideal response trajectory, there were just two principal components, corresponding roughly to horizontal and vertical pieces of the ideal trajectory. Furthermore, the reconstruction of the mean trajectories for different directions combined these components with coefficients that corresponded to the sines and cosines of the relevant directions. Therefore, we were able to identify $\mathbf{v}_{\text{dir}} = \partial \mathbf{v}_{\text{ideal}} / \partial \theta$ with $(\partial \hat{R}(\theta) / \partial \theta)_{\theta=0} \cdot \mathbf{v}_{\text{ideal}}$, where $\hat{R}(\theta)$ is the matrix representing rotation through an angle θ .

To analyse deviations from ideal behaviour on individual trials, we subtracted the mean response for a given target direction from each individual pursuit trial to form a noise vector. We computed the temporal covariance of pursuit noise across all trials (a 250 × 250 matrix), and then subtracted the covariance of the background to form ΔC . We tested alternative noise models to confirm that the low-dimensional structure we observed in ΔC did not arise from our choice of 'background' noise. First, we used a 'white' noise model in which errors were independent in 1-ms bins, and eye-velocity variance grew as a function of the mean eye velocity: ΔC had 80 to 90 significant eigenvalues. Second, we preserved the form of temporal correlations in eye velocity during fixation (Fig. 2a), again with variances that scaled with the mean response: the three dominant eigenvalues captured only 67.5% of the variance and the axes defined by speed, direction and time accounted for less than half of the total variance. Statistical analysis of these models confirmed that the observed low-dimensional structure of trial-by-trial variations in the pursuit trajectory had a very low probability of occurring by chance ($<10^{-5}$).

Received 9 June; accepted 24 June 2005.

1. Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psychol.* **47**, 381–391 (1954), reprinted in *J. Exp. Psychol.* **121**, 262–269 (1992).
2. Gordon, J., Ghilardi, M. F. & Ghez, C. Accuracy of planar reaching movements. I. Independence of direction and extent variability. *Exp. Brain Res.* **99**, 97–111 (1994).
3. Messier, J. & Kalaska, J. F. Comparison of variability of initial kinematics and endpoints of reaching movements. *Exp. Brain Res.* **125**, 129–152 (1999).
4. Harris, C. M. & Wolpert, D. M. Signal-dependent noise determines motor planning. *Nature* **394**, 780–784 (1998).
5. d'Avella, A. & Bizzi, E. Low dimensionality of suraspinal induced force fields. *Proc. Natl Acad. Sci. USA* **95**, 7711–7714 (1998).
6. Santello, M., Flanders, M. & Soechting, J. F. Postural hand strategies for tool use. *J. Neurosci.* **18**, 10105–10115 (1998).
7. Sanger, T. D. Human arm movements described by a low-dimensional superposition of principal components. *J. Neurosci.* **20**, 1066–1072 (2000).
8. Todorov, E. & Jordan, M. I. Optimal feedback control as a theory of motor coordination. *Nature Neurosci.* **5**, 1226–1235 (2002).
9. Donchin, O., Francis, J. T. & Shadmehr, R. Quantifying generalization from trial-by-trial behaviour of adaptive systems that learn with basis functions: theory and experiments in human motor control. *J. Neurosci.* **23**, 9032–9045 (2003).
10. Lisberger, S. G., Morris, E. J. & Tyschen, L. Visual motion processing and sensory-motor integration for smooth pursuit eye movements. *Annu. Rev. Neurosci.* **10**, 97–129 (1987).
11. Lisberger, S. G. & Westbrook, L. E. Properties of visual inputs that initiate horizontal smooth pursuit eye movements in monkeys. *J. Neurosci.* **5**, 1662–1673 (1985).
12. de Bruyn, B. & Orban, G. A. Human velocity and direction discrimination measured with random dot patterns. *Vision Res.* **28**, 1323–1335 (1988).
13. Watamaniuk, S. N. J. & Heinen, S. J. Human smooth pursuit direction discrimination. *Vision Res.* **39**, 59–70 (1999).
14. Kowler, E. & McKee, S. P. Sensitivity of smooth eye movement to small differences in target velocity. *Vision Res.* **27**, 993–1015 (1987).
15. Gegenfurter, K. R., Xing, D., Scott, B. H. & Hawken, M. J. A comparison of pursuit eye movement and perceptual performance in speed discrimination. *J. Vis.* **3**, 865–876 (2003).
16. Purushothaman, G. & Bradley, D. C. Neural population code for fine perceptual decisions in area MT. *Nature Neurosci.* **8**, 99–106 (2005).
17. Liu, J. & Newsome, W. T. Correlation between speed perception and neural activity in the medial temporal visual area. *J. Neurosci.* **25**, 711–722 (2005).
18. Newsome, W. T., Britten, K. H. & Movshon, J. A. Neuronal correlates of a perceptual decision. *Nature* **341**, 52–54 (1989).
19. Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
20. Stone, L. S. & Krauzlis, R. J. Shared motion signals for human perceptual decisions and oculomotor actions. *J. Vis.* **3**, 725–736 (2003).
21. Carpenter, R. H. S. in *Eye Movements: Cognition and Visual Perception* (eds Fisher, D. F., Monty, R. A. & Senders, J. W.) 237–246 (Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1981).
22. Osborne, L. C., Bialek, W. & Lisberger, S. G. Time course of information about motion direction in visual area MT. *J. Neurosci.* **24**, 3210–3222 (2004).
23. Roitman, J. D. & Shadlen, M. N. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* **22**, 9475–9489 (2002).
24. Hanes, D. P. & Schall, J. D. Neural control of voluntary movement initiation. *Science* **274**, 427–430 (1996).
25. Ross, J., Morrone, M. C., Goldberg, M. E. & Burr, D. C. Changes in visual perception at the time of saccades. *Trends Neurosci.* **24**, 316–318 (2001).
26. Fuchs, A. F. & Luschei, E. S. Firing patterns of abducens neurons of alert monkeys in relationship to horizontal eye movement. *J. Neurophysiol.* **33**, 382–392 (1970).
27. Bialek, W. in *Physics of Biomolecules and Cells: Les Houches Session LXXV* (eds Flyvbjerg, H., Julicher, F., Ormos, P. & David, F.) 485–577 (EDP Sciences, Les Ulis and Springer-Verlag, Berlin, 2002).
28. Bialek, W. Physical limits to sensation and perception. *Annu. Rev. Biophys. Chem.* **16**, 455–478 (1987).
29. Barlow, H. B. Critical limiting factors in the design of the eye and visual cortex. *Proc. R. Soc. Lond. B* **212**, 1–34 (1981).

Acknowledgements This work was supported in part by a National Institutes of Health Grant and by the Howard Hughes Medical Institute. We thank S. Tokiyama, E. Montgomery and K. MacLeod for assistance with animal monitoring and maintenance, and S. Ruffner for computer programming. W.B. thanks the Sloan-Swartz Center at UCSF for its hospitality during critical stages of this collaboration.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.C.O. (osborne@phy.ucsf.edu).

WNT7b mediates macrophage-induced programmed cell death in patterning of the vasculature

Ivan B. Lobov^{1*†}, Sujata Rao^{1*}, Thomas J. Carroll^{2†}, Jefferson E. Vallance¹, Masataka Ito³, Jennifer K. Ondr¹, Savita Kurup^{1†}, Donald A. Glass⁴, Millan S. Patel⁴, Weiguo Shu⁵, Edward E. Morrissey⁵, Andrew P. McMahon², Gerard Karsenty⁴ & Richard A. Lang¹

Macrophages have a critical role in inflammatory and immune responses through their ability to recognize and engulf apoptotic cells¹. Here we show that macrophages initiate a cell-death programme in target cells by activating the canonical WNT pathway. We show in mice that macrophage WNT7b is a short-range paracrine signal required for WNT-pathway responses and programmed cell death in the vascular endothelial cells of the temporary hyaloid vessels of the developing eye. These findings indicate that macrophages can use WNT ligands to influence cell-fate decisions—including cell death—in adjacent cells, and raise the possibility that they do so in many different cellular contexts.

In most systems, it has largely been assumed that macrophage involvement in programmed cell death comes after the apoptotic event, and is a response to the presence of membrane-tethered or soluble 'eat-me' signals from dead and dying cells. However, in some circumstances phagocytes actively induce programmed cell death. In mice, macrophages are required for the programmed regression of temporary capillary networks within the developing eye^{2,3}. In the worm *Caenorhabditis elegans*, phagocyte recognition pathways can act as a backup stimulus for the induction of programmed cell death when the autonomous caspase-driven pathway is deficient^{4,5}. These data suggest that there are evolutionarily conserved pathways for phagocyte-induced programmed cell death.

Using ablation and repletion experiments^{2,3}, it has been shown that resident macrophages are necessary and sufficient for the programmed cell death that drives the regression of the temporary ocular capillary networks. To determine whether mice deficient in the lymphomyeloid transcription factor *PU.1*, which lack macrophages⁶, are useful for studying programmed vascular regression, we confirmed that resident ocular macrophages were absent in these animals (Fig. 1a–d) and assessed ocular vessel network regression postnatally, when the hyaloid vessels would normally regress⁷. As expected^{2,3}, the pupillary membrane (on the anterior surface of the lens, Supplementary Fig. 1) and the tunica vasculosa lentis (on the posterior surface of the lens, data not shown), as well as the hyaloid vessels (between the lens and retina, Fig. 1e–h), appeared to persist postnatally in *PU.1*-deficient mice. Counts of vessel numbers at postnatal day 8 (P8) (Fig. 1i) confirmed this, and also revealed that *PU.1* heterozygotes had a partial regression phenotype. TdT-mediated dUTP nick end-labelling (TUNEL) analysis showed that *PU.1*-null mice have a defect in apoptotic cell clearance (data not shown) that precluded the meaningful quantification of apoptotic vascular endothelial cells

(VECs). These data establish that *PU.1*-mutant mice are a useful model for studying macrophage function in programmed vascular regression.

The WNT signalling pathway has a crucial function in developmental cell fate decisions^{8–11} and, when aberrantly activated, in the development of cancer¹². In vertebrates, the canonical WNT response requires a receptor complex comprising the coreceptor low-density lipoprotein receptor-related protein (LRP)5 or LRP6 (ref. 13) and a multiple-pass transmembrane receptor of the Frizzled (FZD) family¹⁴. When activated by a WNT ligand, this complex initiates a cascade of events that culminates in the stabilization of β -catenin, its association with transcription factors of the *Lef/Tcf* family¹⁵ and the regulation of target genes including some that stimulate cell cycle entry⁸. Notably, as indicated by the phenotype of the homozygous-null *Lrp5*^{lacZ/lacZ} mice^{16,17} (Fig. 2a–d, g), the coreceptor gene *Lrp5* is required for the regression of the hyaloid vessels as a consequence of reduced VEC apoptosis¹⁷ (Fig. 2h). As LRP5 may have activities outside the canonical WNT pathway⁹, we assessed the consequences of mutating the *Lef1* gene¹⁸ and again found persistent hyaloid vessels (Fig. 2e–g) and reduced levels of cell death (Fig. 2h). As the WNT pathway is often a stimulus for proliferation, we also performed 5-bromodeoxyuridine (BrdU) labelling and showed that in capillary cells of both the *Lrp5* and *Lef1* mutants, the rate of cell cycle entry is reduced (Fig. 2i). Taken together, these data indicate that canonical WNT signalling is required for VEC cell cycle entry, apoptosis and programmed capillary regression.

To determine which cell type was WNT-responsive, we took advantage of canonical WNT pathway reporter mice (*TOPGAL*) that carry a transgene with *Lef/Tcf* binding sites and a minimal promoter upstream of an open reading frame encoding β -galactosidase (β -gal)¹⁹. Staining of hyaloid vessels from *TOPGAL* hemizygous mice with 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal) revealed β -gal-positive VECs adjacent to the vessel lumen (Fig. 2j). The intensity of X-gal staining in VECs increased from P3 (Fig. 2l) to P5 and P7 (Fig. 2m, k). Macrophages could be identified morphologically (Fig. 2j–m, dashed circles) and did not stain. The X-gal staining pattern in P7 hyaloid vessels from *TOPGAL* and *Lrp5*^{lacZ/lacZ} mice were quite distinct, with *TOPGAL* mice showing sporadic intensely stained VECs (Fig. 2j, k) and the *Lrp5*^{lacZ/lacZ} mice showing weak general staining (Fig. 2n). In compound *Lrp5*^{lacZ/lacZ}; *TOPGAL* mice, the intense sporadic VEC labelling was lost (Fig. 2o), consistent with the expectation that LRP5 is required for WNT signalling. When

¹Division of Developmental Biology, The Children's Hospital Research Foundation and Department of Ophthalmology, University of Cincinnati, Cincinnati, Ohio 45229, USA.

²Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ³Department of Anatomy, National Defense Medical College, Tokorozawa 359-8513, Japan. ⁴Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA. ⁵Department of Medicine and Molecular Cardiology Research Center, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [†]Present addresses: Regeneron Pharmaceuticals, Tarrytown, New York 10591, USA (I.B.L.); University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75390, USA (T.J.C.); Department of Medicine, University of Sydney, New South Wales 2006, Australia (S.K.).

*These authors contributed equally to this work.

combined with the reduced cell death and regression failure of the *Lrp5*- and *Lef1*-mutant mice, these data indicate that the WNT response of VECs is critical for scheduled vascular regression.

One way to explain the requirement for macrophages in scheduled vascular regression would be if they were a source of WNT ligands. To investigate this possibility, we determined whether the macrophage-deficient *PU.1*-null mice showed defects that were characteristic of perturbations in the WNT pathway. BrdU labelling showed that, like the *Lrp5* and *Lef1* mutants (Fig. 2i), *PU.1*-null mice show a reduced rate of S-phase labelling in cells of primary and secondary hyaloid capillary branches (Supplementary Fig. 2a). Crosses between *PU.1* and *TOPGAL* mice produced macrophage-deficient *TOPGAL* animals that showed a 70% reduction in the number of X-gal-stained cells in the hyaloid vessels (Supplementary Fig. 2b). These data suggest that macrophages can stimulate WNT pathway activation in hyaloid VECs.

To determine whether macrophages might express a set of *Wnt* genes associated with vascular development²⁰, we performed reverse transcription–polymerase chain reaction (RT–PCR) analysis on purified populations of hyaloid macrophages gathered from wild-type P5 hyaloid preparations using the laser-capture microdissection technique (Supplementary Fig. 2). This revealed that of a small number of *Wnt* genes expressed in macrophages, *Wnt7b* seemed to be specific for hyaloid macrophages (Supplementary Fig. 2). To investigate further the expression of *Wnt7b*, we performed X-gal staining on hyaloid vessel preparations from *Wnt7b^{+/lacZ}* mice²¹ (Fig. 3a–c), which confirmed that *Wnt7b^{lacZ}* expression was restricted to macrophages. Furthermore, the level of X-gal staining increased dramatically from P1 to P5 (Fig. 3a–c) and correlated with increasing

TOPGAL expression in VECs over the same time interval (Fig. 2k–m). These data suggested that macrophages are required for scheduled vascular regression because they are a source of WNT7b.

To determine whether this was the case, we generated a hypomorphic *Wnt7b^{d1/d1}* gene-targeted allele (Supplementary Fig. 3). It was necessary to generate this allele for analysis as previously generated *Wnt7b* mutations, including the *Wnt7b^{lacZ}* allele used above²¹, result in lethality too early to permit analysis of hyaloid vessel regression. Interestingly, *Wnt7b^{d1/d1}* mice show persistence of the hyaloid vessels (Fig. 3d–h). Immunolabelling of wild-type and *Wnt7b^{d1/d1}* hyaloid vessels for the macrophage marker F4/80 showed that normal numbers of macrophages were present in mutant mice (Fig. 3i, j), eliminating the possibility that hyaloid vessel persistence was a consequence of a lack of macrophages at these sites. It should be noted that *Wnt7b^{d1/d1}* mice also showed levels of apoptosis (Fig. 3k) and BrdU labelling (Fig. 3l) that were much reduced from wild-type levels but comparable to the levels observed in *Lrp5*-mutant mice. As *Wnt7b* expression is restricted to macrophages, these data suggest that WNT7b is a macrophage product critical for inducing cell death in hyaloid VECs.

To determine which FZD receptor might mediate WNT7b signalling, we assessed the activity of FZD3–FZD8 using the *SuperTOPFLASH* WNT reporter cell line¹¹ and a luciferase reporter assay (Supplementary Fig. 4). Although we could not exclude the participation of multiple FZDs, the expression of *Fzd4* in hyaloid capillaries, its activity in mediating WNT7b signalling (Supplementary Fig. 4a, b) and the persistence of hyaloid vessels in *Fzd4*-mutant mice¹¹ all implicated FZD4. For this reason, we also examined the specificity of the interaction using a mixed-cell assay in which we tested the ability

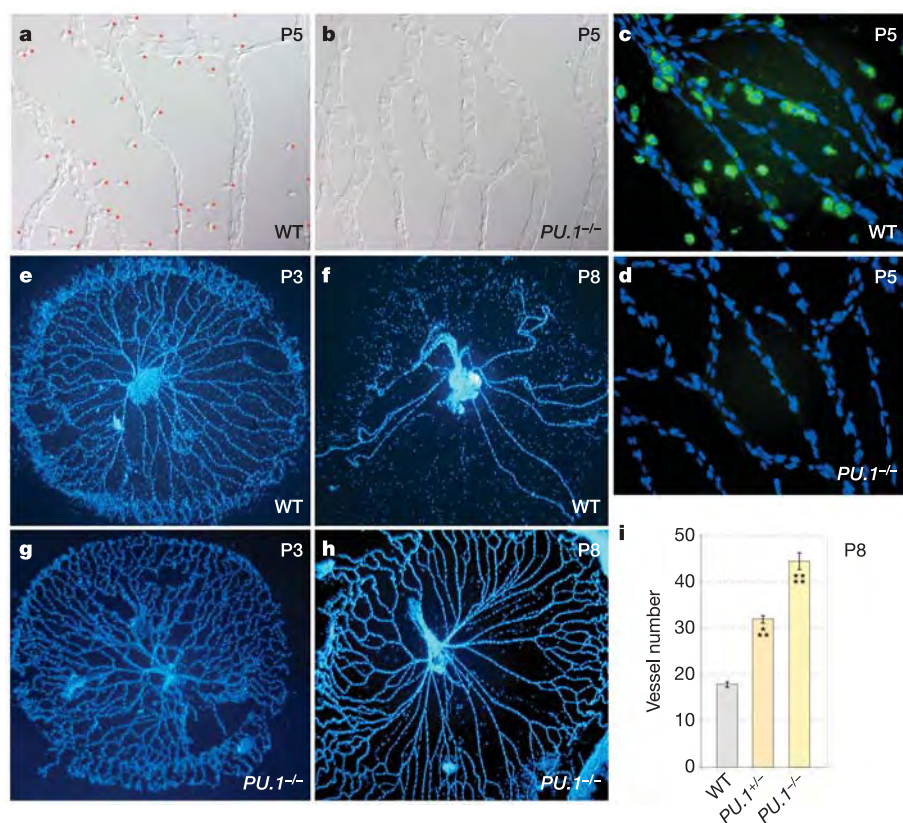


Figure 1 | Regression of the hyaloid vessels is macrophage-dependent. **a–d**, Hyaloid vessel preparations from wild-type (**a**, **c**) and *PU.1*^{-/-} (**b**, **d**) mice at P5. **a**, **b**, Differential interference contrast illumination indicating the presence of macrophages in wild-type mice (**a**, with red dots adjacent) and the absence of macrophages in *PU.1*^{-/-} mice (**b**). **c**, **d**, Fluorescent immunostaining for macrophages (F4/80, a macrophage-specific marker,

green) and nuclei (blue). **e–h**, Hyaloid vessels from wild-type (**e**, **f**) and *PU.1*^{-/-} (**g**, **h**) animals of the indicated ages stained with Hoechst 33258. **i**, Hyaloid vessel number in wild-type, *PU.1*^{+/-} and *PU.1*^{-/-} mice at P8. All error bars are standard errors. Significance levels: three asterisks, 0.0001 < *P* < 0.001; four asterisks, *P* < 0.0001. Original magnification: ×50 (**e–h**); ×400 (**a–d**). WT, wild type.

of FZD4 ligand-binding domain mutants (see Methods) to mediate paracrine signalling by WNT7b (Supplementary Fig. 4c). This showed that subtle mutations in FZD4 prevent paracrine WNT7b signalling, and further support a model in which WNT7b produced by macrophages signals through FZD4 expressed on endothelial cells.

If macrophage WNT7b was signalling to hyaloid VECs and eliciting a WNT pathway response, we would predict that in *TOPGAL;Wnt7b^{d1/d1}* compound mice, the *TOPGAL* response would be reduced. To assess this, we generated X-gal-stained P8 hyaloid vessel preparations from control *TOPGAL* and *TOPGAL;Wnt7b^{d1/d1}* mice (Fig. 4a, b) and counted the numbers of stained cells. The data were normalized to take into account the greater number of vessels in the *Wnt7b^{d1/d1}* mutant; this showed that the *Wnt7b* mutation reduced the *TOPGAL* response to 29% of the control value (Fig. 4c). This corresponds very closely to the reduced *TOPGAL* response observed when macrophages are absent (in *PU.1*-mutant mice, Supplementary Fig. 2b) and further argues that macrophage WNT7b elicits a WNT pathway response in hyaloid VECs.

WNT ligands can be lipid-modified and are therefore highly insoluble²². This suggests that macrophage WNT7b might elicit VEC responses only through cell–cell contact. Therefore, we used a technique described previously¹¹ to assess the range of action of WNT7b using populations of producer and responder cells that were either closely apposed or widely separated in the culture dish. The *SuperTOPFLASH* cell line¹¹ transfected with *Fzd4* and *Lrp5* expression plasmids was used to assess WNT pathway responses. Transfected 293 cells, containing a *Wnt7b* expression plasmid, were used as a producer cell line. Significant activity was detected only when WNT7b-producer cells were mixed with FZD4-expressing responder cells (Fig. 4d), suggesting that WNT7b had a short range of action.

Despite the reduced *TOPGAL* responsiveness in *TOP-*

GAL;Wnt7b^{d1/d1} mice, it remained possible that another source of WNT7b in addition to the macrophage might be critical to initiate the hyaloid VEC death programme. To investigate this possibility, we took advantage of *PU.1*-mutant mice and asked whether injected macrophages could rescue the hyaloid vessel regression failure. Bone-marrow-derived macrophages were differentiated in the presence of CSF-1 (colony-stimulating factor 1) as described previously³ and were allowed to mature to 14 ± 3 days before injection. Intravitreal injection into the right eye was performed on the day of birth; the contralateral eye was used as an uninjected control. As expected, no macrophages were observed in uninjected *PU.1*-mutant eyes and the hyaloid vessels were persistent (Fig. 4e). Injected wild-type macrophages were localized to the hyaloid capillaries (Fig. 4f), and by P8 had rescued regression completely (Fig. 4f, h) according to the number of hyaloid vessels remaining at P8 in wild-type mice (Fig. 3h). Notably, when bone marrow macrophages from *Wnt7b^{d1/d1}* mice were injected, they also localized to hyaloid vessels (Fig. 4g) but completely failed to mediate hyaloid vessel regression (Fig. 4g, h). When combined with the range-of-action experiments and the results from the *TOPGAL;Wnt7b^{d1/d1}* compound mice, these data indicate that the critical, paracrine source of WNT7b required for scheduled vascular regression is the macrophage.

Here we have established that macrophages actively regulate life-and-death decisions in adjacent cells through the WNT pathway. In the context of the hyaloid vessel system, the critical mediator of cell death is WNT7b. Although it is possible that a more complex signalling pattern exists, the data support a model in which macrophage WNT7b activates the WNT pathway in adjacent vascular endothelial cells, most likely through cell–cell contact. As the WNT pathway can stimulate cell cycle entry in this and other systems⁸, there is a possibility that programmed cell death is dependent on cell cycle entry. Coupling of cell death and cell cycle entry has been noted in other systems²³.

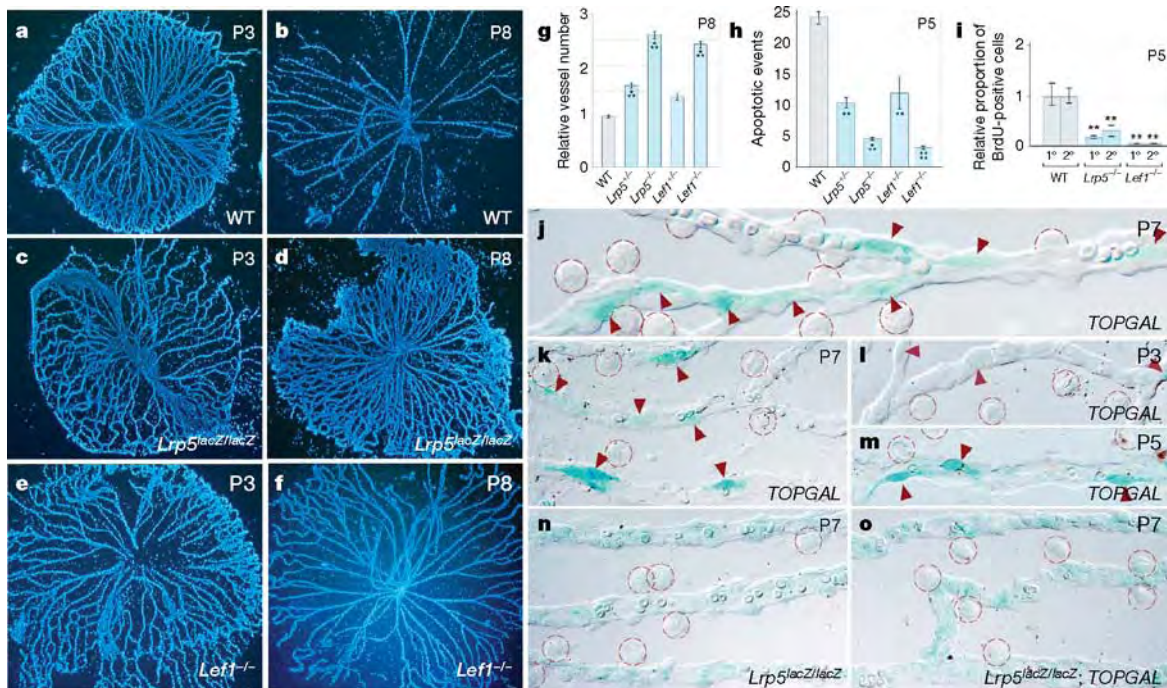


Figure 2 | The WNT pathway response in VECs is required for hyaloid vessel regression. a–f, Hoechst-33258-labelled hyaloid vessel preparations from mice of the indicated ages and genotypes. g–i, Vessel number at P8 (g), number apoptotic events at P5 (h) and BrdU labelling index in primary (1°) and secondary (2°) capillary branches at P5 (i) in the hyaloid vessels of mice with the indicated genotypes. *Lrp5^{lacZ/lacZ}* is an abbreviated reference to the *Lrp5^{lacZ/lacZ}* allele. j–o, X-gal staining (blue) in hyaloid vessels from mice of

the indicated ages and genotypes. Intensely stained *TOPGAL*-expressing cells (red arrowheads) and resident macrophages (red dashed circles) are apparent. All error bars are standard errors. Significance levels: two asterisks, $0.001 < P < 0.01$; three asterisks, $0.0001 < P < 0.001$; four asterisks, $P < 0.0001$. Original magnification: $\times 50$ (a–f); $\times 630$ (k–o); $\times 1,000$ (j). WT, wild type.

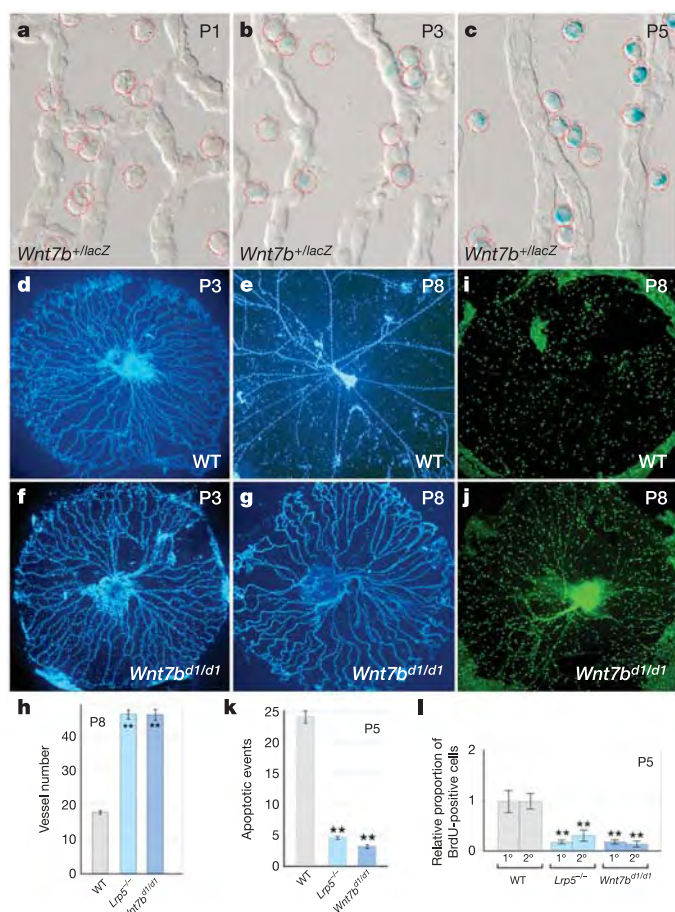


Figure 3 | WNT7b is required for hyaloid vessel regression and is expressed in macrophages. **a–c**, X-gal staining (blue) of hyaloid vessels from *Wnt7b*^{+/lacZ} mice of the indicated ages. Macrophages are indicated by dashed red circles. **d–g**, **i, j**, Hyaloid vessels from wild-type (**d, e, i**) and *Wnt7b*^{d1/d1} (**f, g, j**) animals at the indicated ages, stained with Hoechst 33258 (**d–g**) or anti-F4/80 (**i, j**). **h, k, l**, Quantification, in hyaloid vessels from mice with the indicated genotypes, of vessel number at P8 (**h**), apoptotic events at P5 (**k**) and BrdU labelling index in primary (1°) and secondary (2°) capillary branches at P5 (**l**). All error bars are standard errors. Significance levels: two asterisks, 0.001 < *P* < 0.01. Original magnification: ×50 (**d–g, i, j**); ×630 (**a–c**). WT, wild type.

The current study defines a role for the WNT pathway in regulating vascular regression. Given the broad distribution of macrophages, they may be more generally involved in regulating vascularity through the activation of the WNT pathway in vascular structures^{10,11}. In addition, tumour-associated macrophages are known to express WNT genes²⁴ and can extrinsically regulate the growth of the tumour vasculature, and also perhaps of the tumour cells themselves²⁵. Thus, the finding that macrophages can produce WNT ligands has important implications given the many circumstances in which the WNT pathway has a critical function. The short-range action of macrophage WNT7b is a mechanism for controlling the activity of this potent mediator.

Beyond vascular development, the major question that arises from this analysis is whether the ability of macrophages to actively signal cell death is a unique adaptation in the eye or a more general phenomenon. Although these alternatives need to be tested rigorously, there is some evidence to suggest the latter. For example, in mammalian systems, macrophage-related microglia can promote cell death in the retina through the production of nerve growth factor²⁶. Recently, ablation experiments in a model of liver injury have shown that inflammatory macrophages promote myofibroblast prolifer-

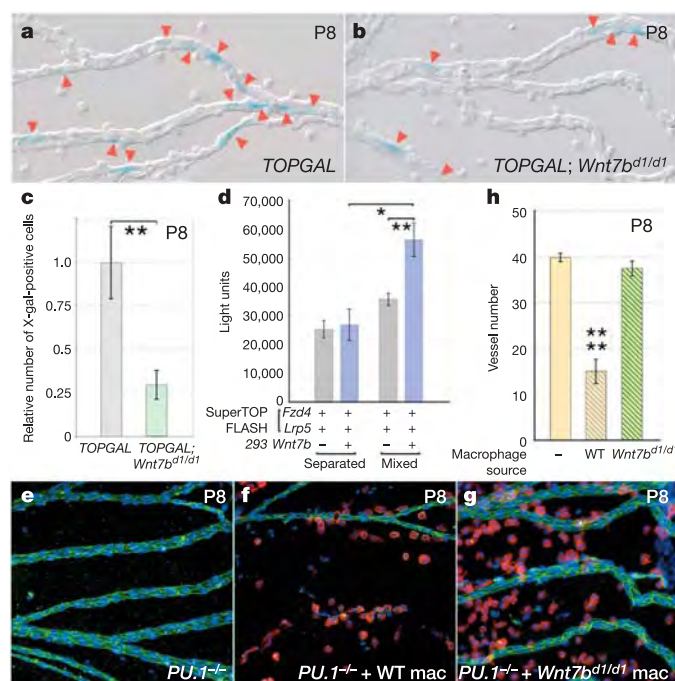


Figure 4 | Macrophages are a critical paracrine source of WNT7b required for hyaloid vessel regression. **a, b**, X-gal staining (blue) of P8 hyaloid vessel preparations in TOPGAL (**a**) and TOPGAL; *Wnt7b*^{d1/d1} (**b**) mice. Red arrowheads indicate stained cells. **c**, Relative number of X-gal-stained cells in TOPGAL (grey bar) and TOPGAL; *Wnt7b*^{d1/d1} (green bar) hyaloid preparations at P8. The data are normalized for the increased vessel number of the *Wnt7b*^{d1/d1} mutants. **d**, Quantification of WNT responsiveness in SuperTOPFLASH cells transfected with *Fzd4* and *Lrp5* and either mixed with WNT7b producer cells or separated from them by restricting SuperTOPFLASH responder cells and WNT7b producer cells to each half of a culture dish. **e–g**, Hyaloid vessel preparations at P8 from uninjected *PU.1*^{-/-} mice (**e**), *PU.1*^{-/-} mice injected with wild-type macrophages (**f**) and *PU.1*^{-/-} mice injected with *Wnt7b*^{d1/d1} macrophages (**g**). Preparations are labelled for the F4/80 marker (red), vascular endothelial cell cadherin (green) and nuclei (blue). Original magnification: ×200. **h**, Vessel number at P8 in uninjected *PU.1*^{-/-} mice (yellow bar), *PU.1*^{-/-} mice injected with wild-type macrophages (yellow/grey bar) and *PU.1*^{-/-} mice injected with *Wnt7b*^{d1/d1} macrophages (yellow/green bar). All error bars are standard errors. Significance levels: one asterisk, 0.01 < *P* < 0.05; two asterisks, 0.001 < *P* < 0.01; four asterisks, *P* < 0.0001. WT, wild type; Mac, macrophages.

ation and apoptosis as a mechanism of resolution and repair²⁷. It has also been shown that phagocytes can promote programmed cell death outside of mammalian systems. In *C. elegans*, the machinery required for the recognition and engulfment of dead cells can promote apoptosis when CED3 (caspase) activity is compromised, suggesting that phagocytes provide a backup mechanism for disposing of superfluous cells^{4,5}. All this would suggest that phagocytes may have a conserved function in signalling cell death in a variety of contexts.

METHODS

Mouse breeding and genotyping. Genotyping of *Lrp5*^{lacZ/lacZ} (ref. 17), *Lef1*^{-/-} (ref. 18), *PU.1*^{-/-} (ref. 6) and TOPGAL (ref. 19) mice was performed as previously described. When *Lef1*-mutant mice were produced as C57BL/6 × 129Sv F₁ hybrids they showed enhanced survival that allowed the analysis of hyaloid vessel regression postnatally.

Dissections, immunostaining and imaging. Hyaloid vessel preparations were generated as previously described¹⁷ with the exception that 5% (w/v) gelatin was used at 56 °C and allowed to set on ice before completion of the dissection. X-gal staining was done according to established protocols²¹. Indirect immunofluorescent staining and BrdU labelling were performed as previously described²⁸. Primary antibodies used were anti-vascular-endothelial-cell-cadherin (Santa

Cruz), anti-BrdU (Dako) and anti-F4/80 (Caltag laboratories), all at a 1:100 dilution. Secondary antibodies labelled with Alexa fluorochromes (Molecular Probes) were used at a 1:500 dilution. TUNEL labelling of apoptotic cells was performed using an *in situ* cell death detection kit (Roche).

Plasmids. All *Fzd*, *Wnt* and *Norrin* complementary DNAs used were mouse. The plasmid encoding the Alkaline phosphatase (AP)–Norrin fusion protein has been described elsewhere¹¹. WNT7b was expressed from pCDNA3.1 as a V5- and His-epitope-tagged fusion protein. In addition, we generated *Fzd4*^{m1} and *Fzd4*^{m2} in pCDNA3.1 using a PCR-based mutagenesis strategy. In *Fzd4*^{m1}, the sequence encoding the cysteine-rich domain (residues 54–QNLGYNV-60) was modified to encode AALAYAA. In *Fzd4*^{m2}, the sequence encoding residues 105–MCT-107 was modified to encode ACA. These alanine mutations in the equivalent regions of murine FZD8 were shown to eliminate binding of *Xenopus* WNT8–AP (ref. 29). These expression plasmids were used in the paracrine and range-of-action signalling assays (Supplementary Fig. 4).

Laser-capture microdissection. To purify hyaloid macrophages, we performed laser-capture microdissection from whole-mount hyaloid vessel preparations using the PixCell II laser-capture microdissection system (Arcturus). The macrophages shown in Supplementary Fig. 2c encircled in red are shown after they have been lifted from the preparation (Supplementary Fig. 2d) on the laser-melted polymer membrane.

Reverse transcription–polymerase chain reaction. RNA from isolated macrophages was purified using PicoPure RNA isolation kit (Arcturus) and subsequent RT–PCR performed using the OneStep RT–PCR kit (Qiagen). Nested primers were used for a second round of PCR-amplification using Takara LA Taq (Takara). Primer nucleotide sequences used were: *Gapdh*, forward (F) 5'-ACT CCACTCACGGCAAATTC-3', reverse (R) 5'-CACATTGGGGGTAGGAACAC-3'; *Wnt2*, F 5'-CGGCCTTTGTTTACGCCATC-3', R 5'-TGAATACAGTAGTCTG GAGAA-3'; *Wnt7b*, F 5'-AAGAACTCCGAGTAGGGAGTCG-3', R 5'-TGCG TTGTACTTCTCCTTGAGC-3'; *Wnt7b*, second round F 5'-CCGAGTAGGG AGTCGAGAGG-3', R 5'-CACACCGTGACACTTACATTCC-3'; *Wnt10b*, F 5'-GTGGTAACGGAAAACCTGAAGC-3', R 5'-CTCATCACACAGCATAA CAGC-3'; *Wnt10b*, second round F 5'-GCCAATTCAGACCTGTGG-3', R 5'-CATAACAGCACCAGTGGAACG-3'; *Fzd4*, F 5'-GCTACAACGTGACCAA GATGC-3', R 5'-CAAACCCAAATCTCTCAGGAC-3'; *Fzd4*, second round F 5'-AACTTAGTGGGACACGAGCTG-3', R 5'-CAGCGTCTCTTGAAGG-3'; *Norrin*, F 5'-GAGAAATCATGTACTAGCTGCATCC-3', R 5'-TGTACCGG TAAGTGGCAGTAAG-3'; *Norrin*, second round F 5'-GGCCATAATGGGAGATA CAGAC-3', R 5'-GAGGACAGTGCTGAAGGACAC-3'; *Lef1*, F 5'-CTACAGCG ACGAGCACTTTC-3', R 5'-AGGATCTGGTTGATAGCTGCAC-3'.

Macrophage rescue. Mouse bone-marrow-derived macrophages were isolated as previously described³. Approximately 4,000 macrophages from wild-type or *Wnt7b*^{d1/d1} animals were injected into the vitreous at P1 in a volume of 200 nl using a modification of the trans-corneal technique described previously³. Animals were killed at P8 and hyaloid vessels dissected.

Luciferase and range-of-action assays. For the Frizzled receptor activity screen, *SuperTOPFLASH* cells¹¹ (which carry a reporter plasmid with *Lef1*/*Tcf* binding sites and a minimal promoter upstream of an open reading frame encoding luciferase) were plated into 60-mm dishes and transfected after 24 h with combinations of 1 µg *Lrp5*–Flag, 1 µg *Wnt7b*–V5His and 0.1 µg *Fzd* plasmid made up to 4 µg with *pIRES*–GFP using Eugene 6 (Roche). Forty-eight hours after transfection, cells were washed with PBS and luciferase activity measured with a luciferase assay kit (Promega). For the mixed-cell paracrine signalling assay, 293 cells were transfected with WNT pathway ligands and overlaid on *SuperTOPFLASH* responder cells that were separately transfected with plasmids encoding LRP5 and a FZD receptor. Luciferase activity was measured as indicated above. WNT7b range-of-action experiments were carried out as described¹¹.

Statistical analysis. Vessel number was quantified using established methods⁷. At least four hyaloid vessel preparations were quantified for each experiment. All data are presented with standard error bars. Student's *t*-test and one-way analysis of variance were used to assess statistical significance.

Received 18 January; accepted 10 June 2005.

- Savill, J., Dransfield, I., Gregory, C. & Haslett, C. A blast from the past: clearance of apoptotic cells regulates immune responses. *Nature Rev. Immunol.* **2**, 965–975 (2002).
- Lang, R. A. & Bishop, M. J. Macrophages are required for cell death and tissue remodeling in the developing mouse eye. *Cell* **74**, 453–462 (1993).
- Diez-Roux, G. & Lang, R. A. Macrophages induce apoptosis in normal cells *in vivo*. *Development* **124**, 3633–3638 (1997).

- Hoepfner, D. J., Hengartner, M. O. & Schnabel, R. Engulfment genes cooperate with *ced-3* to promote cell death in *Caenorhabditis elegans*. *Nature* **412**, 202–206 (2001).
- Reddien, P. W., Cameron, S. & Horvitz, H. R. Phagocytosis promotes programmed cell death in *C. elegans*. *Nature* **412**, 198–202 (2001).
- McKercher, S. R. *et al.* Targeted disruption of the *PU.1* gene results in multiple hematopoietic abnormalities. *EMBO J.* **15**, 5647–5658 (1996).
- Ito, M. & Yoshioka, M. Regression of the hyaloid vessels and pupillary membrane of the mouse. *Anat. Embryol. (Berl.)* **200**, 403–411 (1999).
- Nusse, R. WNT targets. Repression and activation. *Trends Genet.* **15**, 1–3 (1999).
- Huelsken, J. & Birchmeier, W. New aspects of Wnt signalling pathways in higher vertebrates. *Curr. Opin. Genet. Dev.* **11**, 547–553 (2001).
- Ishikawa, T. *et al.* Mouse Wnt receptor gene *Fzd5* is essential for yolk sac and placental angiogenesis. *Development* **128**, 25–33 (2001).
- Xu, Q. *et al.* Vascular development in the retina and inner ear: control by *Norrin* and *Frizzled-4*, a high-affinity ligand–receptor pair. *Cell* **116**, 883–895 (2004).
- Bienz, M. & Clevers, H. Linking colorectal cancer to Wnt signalling. *Cell* **103**, 311–320 (2000).
- He, X., Semenov, M., Tamai, K. & Zeng, X. LDL receptor-related proteins 5 and 6 in Wnt/β-catenin signalling: arrows point the way. *Development* **131**, 1663–1677 (2004).
- Perrimon, N. Serpentine proteins slither into the wingless and hedgehog fields. *Cell* **86**, 513–516 (1996).
- Behrens, J. *et al.* Functional interaction of β-catenin with the transcription factor LEF-1. *Nature* **382**, 638–642 (1996).
- Gong, Y. *et al.* LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. *Cell* **107**, 513–523 (2001).
- Kato, M. *et al.* *Cbfa1*-independent decrease in osteoblast proliferation, osteopenia, and persistent embryonic eye vascularization in mice deficient in *Lrp5*, a Wnt coreceptor. *J. Cell Biol.* **157**, 303–314 (2002).
- van Genderen, C. *et al.* Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. *Genes Dev.* **8**, 2691–2703 (1994).
- DasGupta, R. & Fuchs, E. Multiple roles for activated LEF/TCF transcription complexes during hair follicle development and differentiation. *Development* **126**, 4557–4568 (1999).
- Goodwin, A. M. & D'Amore, P. A. Wnt signalling in the vasculature. *Angiogenesis* **5**, 1–9 (2002).
- Shu, W., Jiang, Y. Q., Lu, M. M. & Morrissey, E. E. Wnt7b regulates mesenchymal proliferation and vascular development in the lung. *Development* **129**, 4831–4842 (2002).
- Willert, K. *et al.* Wnt proteins are lipid-modified and can act as stem cell growth factors. *Nature* **423**, 448–452 (2003).
- Lowe, S. W., Cepero, E. & Evan, G. Intrinsic tumour suppression. *Nature* **432**, 307–315 (2004).
- Smith, K. *et al.* Up-regulation of macrophage *Wnt* gene expression in adenoma–carcinoma progression of human colorectal cancer. *Br. J. Cancer* **81**, 496–502 (1999).
- Pollard, J. W. Tumour-educated macrophages promote tumour progression and metastasis. *Nature Rev. Cancer* **4**, 71–78 (2004).
- Frade, J. M. & Barde, Y. A. Microglia-derived nerve growth factor causes cell death in the developing retina. *Neuron* **20**, 35–41 (1998).
- Duffield, J. S. *et al.* Selective depletion of macrophages reveals distinct, opposing roles during liver injury and repair. *J. Clin. Invest.* **115**, 56–65 (2005).
- Diez-Roux, G., Argilla, M., Makarenkova, H., Ko, K. & Lang, R. A. Macrophages kill capillary cells in G1 phase of the cell cycle during programmed vascular regression. *Development* **126**, 2141–2147 (1999).
- Dann, C. E. *et al.* Insights into Wnt binding and signalling from the structures of two Frizzled cysteine-rich domains. *Nature* **412**, 86–90 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank: P. Speeg for technical assistance; E. Fuchs, L. Niswander and C. Dean for the *TOPGAL* mice; S. McKercher and R. Maki for the *PU.1*-null mice; L. Chan for the *Lrp5*-null mice; and, R. Grosschedl for the *Lef1*-null mice. We are indebted to Q. Xu and J. Nathans for providing the *SuperTOPFLASH* cell line and for the *Fzd* and *Norrin* expression plasmids. This work was supported by NIH grants to E.E.M., A.P.M., G.K. and R.A.L. G.K. was also supported by funds from the March of Dimes, and R.A.L. by funds from the Abrahamson Pediatric Eye Institute Endowment at the Children's Hospital Medical Center of Cincinnati.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.A.L. (Richard.Lang@cchmc.org).

LETTERS

Membrane vesicles traffic signals and facilitate group activities in a prokaryote

Lauren M. Mashburn^{1,2} & Marvin Whiteley^{1,2}

Many bacteria use extracellular signals to communicate and coordinate social activities, a process referred to as quorum sensing¹. Many quorum signals have significant hydrophobic character, and how these signals are trafficked between bacteria within a population is not understood. Here we show that the opportunistic human pathogen *Pseudomonas aeruginosa* packages the signalling molecule 2-heptyl-3-hydroxy-4-quinolone (pseudomonas quinolone signal; PQS)² into membrane vesicles that serve to traffic this molecule within a population. Removal of these vesicles from the bacterial population halts cell–cell communication and inhibits PQS-controlled group behaviour. We also show that PQS actively mediates its own packaging and the packaging of other antimicrobial quinolones produced by *P. aeruginosa* into vesicles. These findings illustrate that a prokaryote possesses a signal trafficking system with features common to those used by higher organisms and outlines a novel mechanism for delivery of a signal critical for coordinating group behaviour in *P. aeruginosa*.

Quorum sensing in the Gram-negative bacterium *P. aeruginosa* involves multiple signals including 3-oxo-dodecanoyl homoserine lactone (3OC12-HSL), butyryl homoserine lactone (C4-HSL) and PQS^{3,4}. These signalling molecules constitute a complex integrated regulatory network that controls the transcription of about 5% of all *P. aeruginosa* genes, including many genes involved in virulence^{4,5}.

The quorum sensing model involves the movement of signals into the extracellular environment; however, some signals such as PQS and 3OC12-HSL have significant hydrophobic character (Table 1), which may hamper their dissemination within a bacterial population. Thus trafficking of these hydrophobic signals within a *P. aeruginosa* population may require specialized processes. *P. aeruginosa* is well known to naturally produce extracellular membrane vesicles (MVs)^{6–8} that could potentially serve as delivery vehicles for these signals in a manner similar to the vesicle trafficking systems used in higher organisms^{9,10}. To examine this, *P. aeruginosa* MVs were isolated, shown to be morphologically similar to previously studied MVs^{6–8} (Supplementary Fig. 1) and tested for the presence of 3OC12-

HSL, C4-HSL and PQS. Thin-layer chromatography (TLC), liquid chromatography–mass spectrometry (LC–MS) and biological assays revealed that *P. aeruginosa* MVs contain about 86% of the PQS produced by *P. aeruginosa*, whereas less than 1% of the 3OC12-HSL and C4-HSL produced by *P. aeruginosa* were present in these vesicles (Table 1, Fig. 1). These results indicate that *P. aeruginosa* packages PQS but not acyl-HSL signalling molecules within MVs.

P. aeruginosa produces at least 55 quinolone/quinolines aside from PQS^{3,11}, including quinolones that possess significant antibiotic activity against Gram-positive bacteria^{3,12}. Because of the structural similarity between these molecules and PQS, we reasoned that antimicrobial quinolones would also be packaged into MVs. To test this, LC–MS along with collision-induced dissociation (CID) were used to examine MVs for the presence of these quinolones. This analysis revealed that *P. aeruginosa* MVs contain other previously identified quinolones as well as PQS (Fig. 2)^{3,11,12}.

P. aeruginosa MVs have been shown to fuse to Gram-negative and Gram-positive bacteria¹³, indicating that PQS-loaded and antimicrobial quinoline-loaded MVs might be biologically active. To evaluate the biological activity of PQS within MVs, a *P. aeruginosa* strain was constructed that contained a deletion in *pqsH*. *pqsH* catalyses the final step in PQS biosynthesis, so the *pqsH* mutant is unable to synthesize PQS. Because PQS regulates the transcription of genes necessary for the production of the secondary phenazine metabolite pyocyanin¹⁴, the *pqsH* mutant produces very low levels

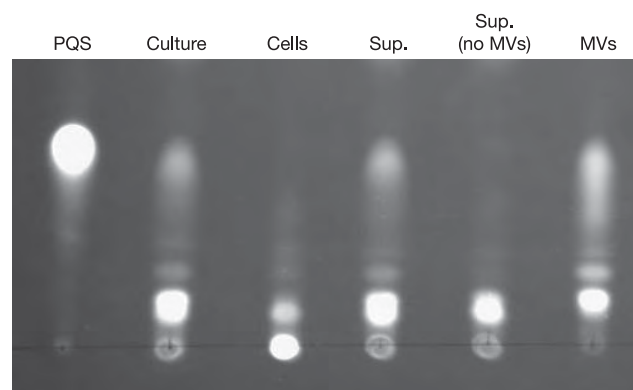


Figure 1 | Packaging of the *P. aeruginosa* signalling molecule PQS into MVs. Culture components were purified as described in Methods, extracted with ethyl acetate and analysed by TLC. Included on the TLC plate are 300 ng of chemically synthesized PQS as a reference (PQS), extract of total culture including cells and supernatant (Culture), extract of bacterial cells from the culture (Cells), extract of cell-free supernatant (Sup.), extract of supernatant after the removal of MVs by ultracentrifugation (Sup. (no MVs)), and extract of purified MVs (MVs).

Table 1 | Localization of *P. aeruginosa* cell-cell signals

Substance	Cells	Percentage of signal in fraction Supernatant (no MVs)	MVs	Hydrophobicity, log <i>P</i>
PQS	9 ± 3	5 ± 5	86 ± 5	3.6
3OC12-HSL	2 ± 1	97 ± 5	<1	1.9
C4-HSL	<1	>99	<1	0.3

Results for percentages of signal are means ± s.d. for six replicate experiments of stationary-phase bacteria (A_{600} of 4.0–7.0) by TLC and *E. coli* bioassays (see Methods). PQS levels were also verified with LC–MS. Hydrophobicity values are predicted log *P* values calculated with the clogP software program. The US Pharmacopeia defines water solubility as log *P* ≤ 0.5.

¹Department of Periodontics, and ²Department of Microbiology and Immunology, The University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104, USA.

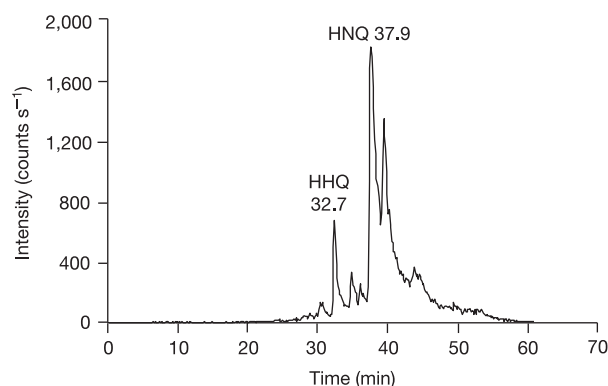


Figure 2 | Analysis of antimicrobial quinolones from *P. aeruginosa* MVs using LC-MS and CID. Shown is an ion chromatogram for quinolones within MVs, which present a major fragment ion at m/z 172. Listed on the elution profile are two quinolones, namely 4-hydroxy-2-heptylquinoline (HHQ) and 4-hydroxy-2-nonylquinoline (HNQ), and their elution times. $[M + H]^+$ ions for HHQ and HNQ had relative masses of 244 and 272, respectively.

of pyocyanin (Fig. 3a); however, the exogenous addition of chemically synthesized PQS to this strain stimulates pyocyanin production (Fig. 3a). The addition of physiological concentrations of MVs to this mutant also increased pyocyanin levels, indicating that PQS contained within MVs is biologically active (Fig. 3a). Organic extraction of PQS from MVs also enhanced the production of pyocyanin, indicating that the inclusion of PQS within MVs is not required for biological activity (Fig. 3a). As expected, pyocyanin production was not stimulated by the addition of MV-free supernatant extracts to the *pqsH* mutant, indicating that the PQS biological activity is associated with MVs.

We next examined the biological activity of the antimicrobial quinolones present within the MVs by evaluating the ability of *P. aeruginosa* MVs to inhibit the growth of actively dividing cells of

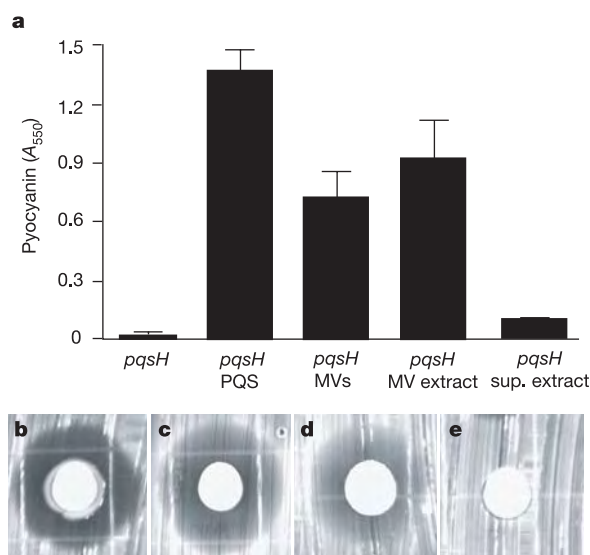


Figure 3 | Biological activities of MVs. **a**, Pyocyanin levels produced by the *pqsH* mutant in the presence of synthetic PQS (25 μ M), MVs (containing 15–25 μ M PQS), ethyl-acetate-extracted MVs and ethyl-acetate-extracted MV-free supernatant. Error bars represent s.d. **b–e**, Antimicrobial activity of MVs from *P. aeruginosa*. A BHI agar Petri dish was swabbed with a confluent lawn of *S. epidermidis* and sterile discs were placed on the agar surface. Discs were then inoculated with an overnight culture of *P. aeruginosa* (**b**), purified MVs (**c**), ethyl acetate extract of purified MVs (**d**) or ethyl acetate as a control (**e**).

the Gram-positive bacterium *Staphylococcus epidermidis*. Application of MVs directly to a lawn of *S. epidermidis* cells produced a zone of clearing similar to that produced by actively growing cells of *P. aeruginosa* (Fig. 3a, b). These results indicate that *P. aeruginosa* MVs possess significant antimicrobial activity. MVs from *P. aeruginosa* have previously been shown to lyse non-dividing Gram-positive bacteria, but this activity was attributed to proteins within the MVs^{13,15}. To determine whether the lysis observed by MVs was due to proteins or antimicrobial quinolones within the MVs, we extracted the quinolones from the MVs by using an organic solvent. Over 80% of the MV antibiotic activity was extractable with this solvent, indicating that most of the antimicrobial activity was due to quinolones and not proteins within the MVs (Fig. 3c–e).

Do the quinolone/quinolones produced by *P. aeruginosa* actively initiate MV formation? To test this, MV formation was examined in a *P. aeruginosa* strain that possesses a mutation in the *pqsA* gene. PqsA is required for the production of all known quinolone/quinolones in *P. aeruginosa*. MV production by the *pqsA* mutant was nearly abolished (Fig. 4a), indicating that quinoline/quinolone production is crucial for MV formation in *P. aeruginosa*.

Because the *pqsA* mutant is deficient in the production of all quinolones/quinolones, it was not clear whether all or a subset of these molecules is required for MV formation. To address this question, we evaluated MV formation by the *pqsH* mutant. Although the *pqsH* mutant does not synthesize PQS, all other classes of quinolones are produced by this strain. The *pqsH* mutant was also highly deficient in MV formation; however, the addition of physiological levels of chemically synthesized PQS to cultures of this mutant

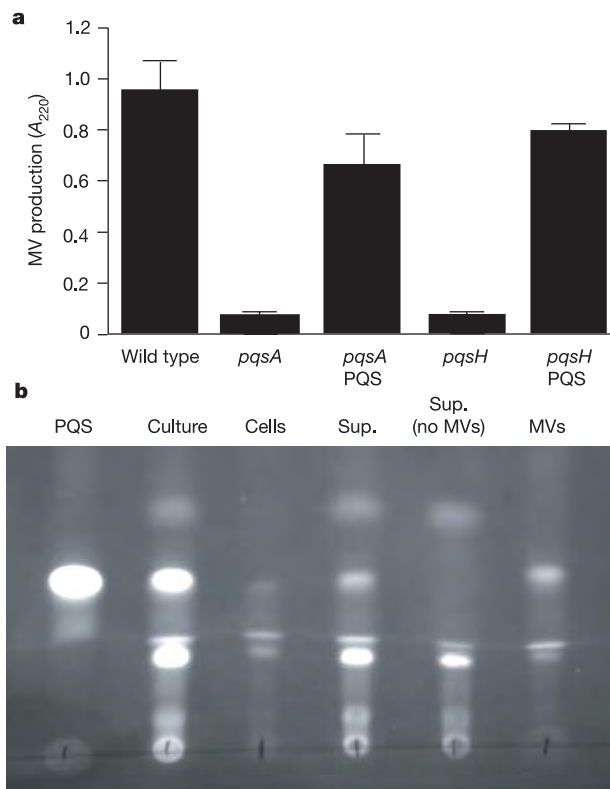


Figure 4 | Exogenous PQS stimulates *P. aeruginosa* MV formation and is packaged into MVs. **a**, MV formation by the *pqsA* mutant and the *pqsH* mutant with and without 50 μ M PQS. Error bars represent s.d. **b**, Examination of the localization of PQS after exogenous addition to the *pqsH* mutant. Culture components were analysed by TLC. Included on the TLC plate were one of the following: 300 ng of chemically synthesized PQS as a reference (PQS), extract of total culture (Culture), extract of bacterial cells (Cells), extract of cell-free supernatant (Sup.), extract of supernatant after the removal of MVs (Sup. (no MVs)), and extract of purified MVs (MVs).

restored MV formation (Fig. 4a). In fact, not only did PQS addition stimulate MV formation by PA14-*pqsH*, but 60–80% of the added PQS was packaged within these MVs (Fig. 4b). The addition of synthetic PQS to the *pqsA* mutant also stimulated MV formation (Fig. 4b), indicating that PQS alone might be sufficient to induce MV formation. Furthermore, MVs produced by the *pqsH* mutant after the addition of PQS had significant antimicrobial activity, indicating that antimicrobial quinolones were also packaged within these MVs (data not shown). Taken together, these data indicate that PQS is required and sufficient for MV formation in *P. aeruginosa* and that exogenous PQS mediates its own packaging and the packaging of other quinolones into these vesicles.

PQS signalling controls the transcription of numerous genes in *P. aeruginosa*^{14,16}, so it was not clear whether PQS or factors regulated by PQS are important for the production of MVs. To address this question, we evaluated the ability of exogenously added PQS to stimulate MV formation by the *pqsH* mutant in the presence of the protein synthesis inhibitor tetracycline. The addition of PQS stimulated MV formation in tetracycline-treated *pqsH* mutant bacteria (Fig. 5), indicating that active growth and protein synthesis are not required for MV stimulation by PQS. To confirm this observation, MV formation was evaluated with a *P. aeruginosa* strain that does not produce the PQS-dependent transcriptional regulator MvfR^{16,17}. MvfR is required for PQS production and PQS-mediated gene regulation; thus, the *mvfR* mutant is unable to induce the transcription of PQS-controlled genes even in the presence of exogenous PQS. As expected, the *mvfR* mutant did not produce MVs; however, the exogenous addition of PQS stimulated MV formation in this strain (Fig. 5). These data indicate that PQS-mediated MV formation in *P. aeruginosa* is not due to the regulatory role of PQS but instead implicates a direct role for this molecule in MV formation and signal trafficking. Although several possibilities exist, one potential mechanism for PQS-mediated MV formation is by means of PQS interaction with, and destabilization of, the outer membrane^{8,18}.

The advantage of packaging quinolones/quinolines into MVs might serve several functions. In some environments, dissemination of these hydrophobic molecules might be facilitated by packaging into MVs. Because *P. aeruginosa* MVs fuse with and deliver internal contents to Gram-negative and Gram-positive bacteria¹³, MVs might act to concentrate quinolone/quinolines to allow their mass delivery to surrounding bacteria. Recent studies also indicate that acyl-HSL signals are degraded by prokaryotic and eukaryotic cells, and this

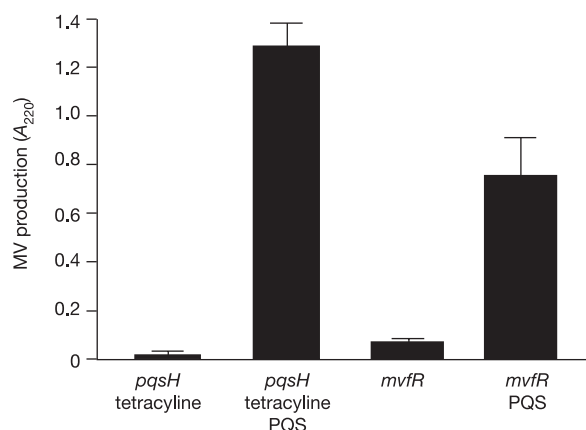


Figure 5 | PQS-mediated gene regulation is not required for MV formation. MV formation was assessed for protein-synthesis-inhibited (by the antibiotic tetracycline) *pqsH* mutant cells in the presence and absence of exogenous PQS. Similar results were obtained with the protein synthesis inhibitor chloramphenicol (data not shown), and the addition of protein synthesis inhibitors halted the growth of *P. aeruginosa*. MV formation was also assessed for the *mvfR* mutant in the presence and absence of exogenous PQS. Error bars represent s.d.

degradation can halt communication within a bacterial population^{19–22}. The chemical structure of PQS makes it a likely substrate for such degradation, so it is plausible that packaging into MVs might protect PQS from catabolism by surrounding cells. These studies also have implications in medicine. *P. aeruginosa* colonizes the lungs of most patients with cystic fibrosis and these infections are often maintained throughout the life of the patient. Cell–cell signalling and antimicrobial quinolone production by *P. aeruginosa* have been proposed to be important for virulence, antibiotic resistance and competition with other bacteria in the lung of patients with cystic fibrosis^{1,12,23}, indicating that inhibiting PQS-mediated MV formation could aid in the treatment of *P. aeruginosa* infections in cystic fibrosis.

Our study indicates that a bacterial cell–cell signal mediates the packaging of itself and other small molecules into MVs. The use of MVs to coordinate group behaviour in a prokaryote draws parallels to eukaryotic vesicle trafficking systems and illustrates that prokaryotes possess communication systems analogous to those commonly used by multicellular organisms. As in higher organisms, vesicle trafficking in *P. aeruginosa* is an active process that requires specific components, and inhibition of this process significantly alters the physiology of the cell population. These findings also provide evidence for the physiological relevance of bacterial MVs in cell–cell signalling and significantly enhance our understanding of the regulation of prokaryotic social activities.

METHODS

Bacterial strains. *P. aeruginosa* strain PA14 and the *pqsA* mutant were obtained from a publicly available database (<http://pga.mgh.harvard.edu/cgi-bin/pa14/mutants/retrieve.cgi>) and the *mvfR* mutant was described previously¹⁷. The *pqsH* mutant was constructed by double homologous recombination as described previously²⁴, resulting in an insertion of *aacC1* (encoding gentamicin resistance) between base pairs 358 and 359 of the *pqsH* coding region. Brain heart infusion (BHI) agar or MOPS medium containing 20 mM succinate²⁵ was used as growth medium. When added to medium, PQS or ethyl acetate extracts were evaporated under a continuous stream of N₂ before resuspension in medium.

Isolation and quantification of MVs. For MV preparation, an overnight culture of *P. aeruginosa* was diluted to 10³–10⁴ bacteria ml^{−1} in BHI broth and allowed to grow at 37 °C for 16–18 h, with shaking at 250 r.p.m. MVs were purified as described previously⁶, except that ultracentrifugation was performed at 265,000 g for 60 min in a Beckman TLA 100.3 rotor. For quantitative purposes, MV preparations were subjected to one buffer exchange with Nanosep 300-kDa omega centrifugal devices (Pall Life Sciences). Protein content of the MVs, which has been previously used to standardize MV preparations⁶, was used as a measure of MV levels by evaluating the absorbance at 220 nm. In some cases, 50 μM synthetic PQS was added to BHI at the beginning of the experiment.

TLC and acyl-HSL bioassays. TLC was performed as described previously⁷, with a 17:2:1 dichloromethane:acetonitrile:dioxane solvent system. One-ml volumes of total bacterial culture, bacterial cells alone, cell-free supernatant (filtered through 0.45-μm and 0.22-μm filters), ultracentrifuged supernatant (without MVs) and MVs were extracted with 3 ml of acidified ethyl acetate and the solvent was evaporated under a continuous stream of N₂. The extracts were concentrated 2–6-fold by resuspension in a 1:1 mixture of ethyl acetate and acetonitrile; 10 μl of this was analysed by TLC. TLC plates were analysed under ultraviolet excitation with an Alpha Innotech FluorChem 8900 imaging system. Acyl-HSLs were quantified with the use of standard *Escherichia coli* bioassays, with 3OC12-HSL and C4-HSL as standards²⁴.

LC–MS and CID. Purified MVs were extracted with three volumes of acidified ethyl acetate, the solvent was evaporated under a continuous stream of N₂, and the extract was resuspended in 30% acetonitrile, 1% acetic acid. This extract was applied to a Magic 2002 HPLC system (Michrom Bioresources, Inc.) with a 5-μm, 1-mm × 150-mm Higgins Cliepus C₈ column connected to a Q-Star mass spectrometer (Applied Biosystems). Elution was with a gradient of 30–100% acetonitrile in water containing 1% acetic acid over a 60-min period at a flow rate of 40 μl min^{−1}. Quinolones were quantified with a precursor ion scan of 172 (ref. 11).

Transmission electron microscopy. MVs were placed on Formvar-coated nickel grids and stained for 10 s with 2% phosphotungstic acid. Samples were viewed with a Hitachi H-7600 transmission electron microscope at 80 kV.

Biological assays. Pyocyanin assays were performed from *P. aeruginosa* grown in MOPS succinate medium. One-ml equivalent volumes of purified MVs, ethyl acetate extracts of purified MVs or ethyl acetate extracts of MV-free supernatants

were added to 1 ml of MOPS succinate medium. Overnight cultures of *P. aeruginosa* were diluted 1:1000 into these media and allowed to grow for 24 h at 37 °C, with shaking. Pyocyanin was extracted from these cultures and quantified spectrophotometrically at 520 nm as described previously²⁶.

Lysis of *S. epidermidis* on Petri dishes was performed by thoroughly swabbing a BHI plate with an overnight culture of *S. epidermidis* diluted to a A_{600} of 0.1–0.2. After drying, sterile test discs (6.4 mm in diameter; Schleicher and Schuell) were placed on the surface of the agar. To these discs was added one of the following: 5 µl of an overnight culture of *P. aeruginosa*; 20 µl of MVs (sixfold concentrate); 20 µl of ethyl acetate extract of MVs (extract was from 20 µl of sixfold concentrated MVs); or 20 µl of ethyl acetate as a control. Discs were allowed to dry at 25 °C and then Petri dishes were incubated at 37 °C for 24 h. Plates were imaged with an Alpha Innotech FluorChem 8900 imaging system and zones of inhibition were measured.

Inhibition of protein synthesis. For protein synthesis inhibition experiments, 25 µg ml⁻¹ tetracycline was added to a mid-exponential phase (A_{600} of 0.3–0.4) *pqsH* mutant culture and incubated for 1 h at 37 °C. The culture was then split into two cultures; one received no treatment and the other received 50 µM PQS. After 4 h, MVs were purified from these cultures and quantified as outlined above.

Received 13 April; accepted 14 June 2005.

- Parsek, M. R. & Greenberg, E. P. Acyl-homoserine lactone quorum sensing in Gram-negative bacteria: A signalling mechanism involved in associations with higher organisms. *Proc. Natl Acad. Sci. USA* **97**, 8789–8793 (2000).
- Pesci, E. C. *et al.* Quinolone signalling in the cell-to-cell communication system of *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **96**, 11229–11234 (1999).
- Deziel, E. *et al.* Analysis of *Pseudomonas aeruginosa* 4-hydroxy-2-alkylquinolines (HAQs) reveals a role for 4-hydroxy-2-heptylquinoline in cell-to-cell communication. *Proc. Natl Acad. Sci. USA* **101**, 1339–1344 (2004).
- Wagner, V. E., Bushnell, D., Passador, L., Brooks, A. I. & Iglewski, B. H. Microarray analysis of *Pseudomonas aeruginosa* quorum-sensing regulons: effects of growth phase and environment. *J. Bacteriol.* **185**, 2080–2095 (2003).
- Schuster, M., Lostroh, C. P., Ogi, T. & Greenberg, E. P. Identification, timing, and signal specificity of *Pseudomonas aeruginosa* quorum-controlled genes: a transcriptome analysis. *J. Bacteriol.* **185**, 2066–2079 (2003).
- Kadurugamuwa, J. L. & Beveridge, T. J. Virulence factors are released from *Pseudomonas aeruginosa* in association with membrane vesicles during normal growth and exposure to gentamicin: a novel mechanism of enzyme secretion. *J. Bacteriol.* **177**, 3998–4008 (1995).
- Nguyen, T. T., Saxena, A. & Beveridge, T. J. Effect of surface lipopolysaccharide on the nature of membrane vesicles liberated from the Gram-negative bacterium *Pseudomonas aeruginosa*. *J. Electron Microsc. (Tokyo)* **52**, 465–469 (2003).
- Kadurugamuwa, J. L. & Beveridge, T. J. Natural release of virulence factors in membrane vesicles by *Pseudomonas aeruginosa* and the effect of aminoglycoside antibiotics on their release. *J. Antimicrob. Chemother.* **40**, 615–621 (1997).
- Rodriguez-Boulan, E., Kreitzer, G. & Musch, A. Organization of vesicular trafficking in epithelia. *Nature Rev. Mol. Cell Biol.* **6**, 233–247 (2005).
- Gallagher, K. L. & Benfey, P. N. Not just another hole in the wall: understanding intercellular protein trafficking. *Genes Dev.* **19**, 189–195 (2005).
- Lepine, F., Milot, S., Deziel, E., He, J. & Rahme, L. G. Electrospray/mass spectrometric identification and analysis of 4-hydroxy-2-alkylquinolines (HAQs) produced by *Pseudomonas aeruginosa*. *J. Am. Soc. Mass Spectrom.* **15**, 862–869 (2004).
- Machan, Z. A., Taylor, G. W., Pitt, T. L., Cole, P. J. & Wilson, R. 2-Heptyl-4-hydroxyquinoline N-oxide, an antistaphylococcal agent produced by *Pseudomonas aeruginosa*. *J. Antimicrob. Chemother.* **30**, 615–623 (1992).
- Kadurugamuwa, J. L. & Beveridge, T. J. Bacteriolytic effect of membrane vesicles from *Pseudomonas aeruginosa* on other bacteria including pathogens: conceptually new antibiotics. *J. Bacteriol.* **178**, 2767–2774 (1996).
- Gallagher, L. A., McKnight, S. L., Kuznetsova, M. S., Pesci, E. C. & Manoil, C. Functions required for extracellular quinolone signalling by *Pseudomonas aeruginosa*. *J. Bacteriol.* **184**, 6472–6480 (2002).
- Kadurugamuwa, J. L. *et al.* S-layered *Aneurinibacillus* and *Bacillus* spp. are susceptible to the lytic action of *Pseudomonas aeruginosa* membrane vesicles. *J. Bacteriol.* **180**, 2306–2311 (1998).
- Deziel, E. *et al.* The contribution of MvfR to *Pseudomonas aeruginosa* pathogenesis and quorum sensing circuitry regulation: multiple quorum sensing-regulated genes are modulated without affecting *lasRI*, *rhlRI* or the production of N-acyl-L-homoserine lactones. *Mol. Microbiol.* **55**, 998–1014 (2005).
- Cao, H. *et al.* A quorum sensing-associated virulence gene of *Pseudomonas aeruginosa* encodes a LysR-like transcription regulator with a unique self-regulatory mechanism. *Proc. Natl Acad. Sci. USA* **98**, 14613–14618 (2001).
- Kadurugamuwa, J. L., Clarke, A. J. & Beveridge, T. J. Surface action of gentamicin on *Pseudomonas aeruginosa*. *J. Bacteriol.* **175**, 5798–5805 (1993).
- Lin, Y. H. *et al.* Acyl-homoserine lactone acylase from *Ralstonia* strain XJ12B represents a novel and potent class of quorum-quenching enzymes. *Mol. Microbiol.* **47**, 849–860 (2003).
- Wang, Y. J. & Leadbetter, J. R. Rapid acyl-homoserine lactone quorum signal biodegradation in diverse soils. *Appl. Environ. Microbiol.* **71**, 1291–1299 (2005).
- Leadbetter, J. R. & Greenberg, E. P. Metabolism of acyl-homoserine lactone quorum-sensing signals by *Variovorax paradoxus*. *J. Bacteriol.* **182**, 6921–6926 (2000).
- Chun, C. K., Ozer, E. A., Welsh, M. J., Zabner, J. & Greenberg, E. P. Inactivation of a *Pseudomonas aeruginosa* quorum-sensing signal by human airway epithelia. *Proc. Natl Acad. Sci. USA* **101**, 3587–3590 (2004).
- Ciofu, O., Beveridge, T. J., Kadurugamuwa, J., Walther-Rasmussen, J. & Hoiby, N. Chromosomal beta-lactamase is packaged into membrane vesicles and secreted from *Pseudomonas aeruginosa*. *J. Antimicrob. Chemother.* **45**, 9–13 (2000).
- Whiteley, M., Parsek, M. R. & Greenberg, E. P. Regulation of quorum sensing by RpoS in *Pseudomonas aeruginosa*. *J. Bacteriol.* **182**, 4356–4360 (2000).
- Whiteley, M. *et al.* Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature* **413**, 860–864 (2001).
- Essar, D. W., Eberly, L., Hadero, A. & Crawford, I. P. Identification and characterization of genes for a second anthranilate synthase in *Pseudomonas aeruginosa*: interchangeability of the two anthranilate synthases and evolutionary implications. *J. Bacteriol.* **172**, 884–900 (1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Jackson for LC-MS analysis; L. Rahme for the *mvfR* mutant; and the MGH-Parabiosys:NHLBI Program for Genomic applications, Massachusetts General Hospital and Harvard Medical School (<http://pga.mgh.harvard.edu/cgi-bin/pa14/mutants/retrieve.cgi>), for the *pqsA* mutant. This work was supported by a grant from the Oklahoma Center for the Advancement of Science and Technology.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.W. (marvin-whiteley@ouhsc.edu).

LETTERS

A mechanosensory complex that mediates the endothelial cell response to fluid shear stress

Eleni Tzima^{1†}, Mohamed Irani-Tehrani¹, William B. Kiosses¹, Elizabetta Dejana², David A. Schultz³, Britta Engelhardt⁴, Gaoyuan Cao⁵, Horace DeLisser⁵ & Martin Alexander Schwartz^{1,6}

Shear stress is a fundamental determinant of vascular homeostasis, regulating vascular remodelling, cardiac development and atherogenesis¹, but the mechanisms of transduction are poorly understood. Previous work showed that the conversion of integrins to a high-affinity state mediates a subset of shear responses, including cell alignment and gene expression^{2–4}. Here we investigate the pathway upstream of integrin activation. PECAM-1 (which directly transmits mechanical force), vascular endothelial cell cadherin (which functions as an adaptor) and VEGFR2 (which activates phosphatidylinositol-3-OH kinase) comprise a mechanosensory complex. Together, these receptors are sufficient to confer responsiveness to flow in heterologous cells. In support of the relevance of this pathway *in vivo*, PECAM-1-knockout mice do not activate NF- κ B and downstream inflammatory genes in regions of disturbed flow. Therefore, this mechanosensing pathway is required for the earliest-known events in atherogenesis.

Atherosclerotic lesions occur preferentially in regions of low or disturbed shear stress at vessel branch points, bifurcations and regions of high curvature, whereas high laminar shear stress is atheroprotective¹. It has been proposed that shear stress is transmitted from the apical surface of the endothelial cell through the cytoskeleton to points of attachment at cell–cell and cell–matrix adhesions; therefore, these adherens junctions experience changes in mechanical tension and could serve as mechanotransducers¹. The endothelium contains adherens junctions that depend on vascular endothelial cell cadherin (VE-cadherin), and tight junctions involving claudins and occludins. Endothelial cells also contain three widely expressed junctional adhesion molecules, and the immunoglobulin family receptor platelet endothelial cell adhesion molecule (PECAM)-1. PECAM-1 is a homophilic adhesion receptor whose cytoplasmic domain binds to β - and γ -catenins and contains an immunoreceptor tyrosine-based inhibitory motif (ITIM) that, in its phosphorylated form, binds to SH2-domain-containing protein tyrosine phosphatase (SHP)-2 (ref. 5).

Acute onset of shear stress triggers a number of events¹, including the activation or induction of ion channels, Src-family and vascular endothelial growth factor receptor 2 (VEGFR2) tyrosine kinases, extracellular signal-regulated kinases (ERKs), c-Jun amino-terminal kinases (JNKs), p38 mitogen-activated protein kinase and AKT (v-akt murine thymoma viral oncogene homologue) serine/threonine kinases, and transcription factors such as NF- κ B (nuclear factor of kappa light chain gene enhancer in B cells) and activator protein (AP)-1. VE-cadherin is implicated in activation of AKT kinases and NF- κ B by flow⁶. The onset of flow also triggers phosphorylation of

PECAM-1 ITIM tyrosines, and direct application of force to PECAM-1 can induce ERK activation⁷. In laminar shear, vascular endothelial cells adapt to the flow so that after the initial stimulation these events are downregulated. In contrast, in disturbed shear, where flow magnitude and direction are continually changing, these pathways are activated in a sustained manner⁸. As a result, NF- κ B is activated and NF- κ B-dependent genes, encoding proteins such as intercellular adhesion molecule (ICAM)-1, vascular cell adhesion molecule (VCAM)-1, endothelial (E)-selectin and platelet-derived growth factor (PDGF), are expressed at atherosclerosis-prone sites *in vivo* before, or in the absence of, other markers for atherosclerosis⁹. Stimulation of these pathways by changes in flow is thought to contribute to the initiation of atherosclerosis, so the acute onset of laminar shear is commonly used to investigate fundamental mechanisms of mechanotransduction.

Although many of the signalling events stimulated by shear have been identified, an understanding of shear sensing has not emerged and the primary transducer(s) that mediate both adaptive responses and atherogenesis have not been identified. The conformational activation of integrins initiates both alignment (an adaptive response to laminar shear) and activation of NF- κ B (which promotes atherogenesis in disturbed shear^{2,3}), so we investigated how the onset of flow induces integrin activation.

Phosphatidylinositol-3-OH kinase (PI(3)K) has been implicated in integrin activation in many cell types¹⁰. In endothelial cells, phosphorylation of the p85 subunit of PI(3)K was detected within 15 seconds after the onset of flow, and continued to increase for several minutes (Supplementary Fig. S1a). When phosphatidylinositol-3,4,5-trisphosphate, the product of PI(3)K, was assayed by examining the membrane translocation of the AKT pleckstrin homology domain fused to green fluorescent protein (GFP-AKT PH), rapid activation by the onset of flow was observed (Supplementary Fig. S1b). When the activation of integrin $\alpha_v\beta_3$ by shear stress was assayed by binding of WOW-1, a monoclonal antibody Fab fragment that binds selectively to high-affinity α_v integrins^{2,11}, a complete block in integrin activation was observed in the presence of the PI(3)K inhibitors LY294002 (Supplementary Fig. S1c) and wortmannin (data not shown). Thus, the stimulation of PI(3)K mediates integrin activation.

Shear stress is known to activate c-Src within seconds^{12,13}. Antibodies against both activating (Y418) and inhibitory (Y527) phosphorylation sites showed that Src was maximally activated within 15 s of the initiation of flow (Supplementary Fig. S1d); it is thus slightly faster than PI(3)K phosphorylation. Both the phosphorylation of the

¹Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. ²Mario Negri Institute for Pharmacological Research and FIRC Institute of Molecular Oncology, Department of Biomolecular and Biotechnological Sciences, Faculty of Sciences, University of Milan, 20139 Milan, Italy. ³Department of Physics, University of California at San Diego, La Jolla, California 92093, USA. ⁴Theodor Kocher Institute, University of Bern, Freiestr. 1, CH-3012 Bern, Switzerland. ⁵Pulmonary and Critical Care Division, Department of Medicine, University of Pennsylvania Medical Center, Philadelphia, Pennsylvania 19104, USA. ⁶Departments of Microbiology and Biomedical Engineering, Cardiovascular Research Center, Mellon Prostate Cancer Research Center, University of Virginia, 415 Lane Road, Charlottesville, Virginia 22908, USA.

[†]Present address: Department of Cell and Molecular Physiology, University of North Carolina at Chapel Hill, 6312 Medical Biomolecular Research Building, 103 Mason Farm Road, Chapel Hill, North Carolina 27599, USA.

PI(3)K p85 subunit and integrin activation were prevented by the Src kinase inhibitors PP2 (Supplementary Fig. S1e) or SU6656 (data not shown). Therefore, a Src family kinase is upstream of PI(3)K-dependent integrin activation in response to shear stress.

To address the requirement for junctional receptors in flow-induced integrin activation, VE-cadherin^{-/-} and PECAM-1^{-/-} endothelial cells were tested for WOW-1 binding. Responses in reconstituted null cells, in which VE-cadherin or PECAM-1 was ectopically re-expressed, were similar to those in wild-type bovine and human endothelial cells (data not shown). VE-cadherin^{-/-} and PECAM-1^{-/-} endothelial cells failed to show strong activation of integrins after the onset of flow, whereas null cells that had been engineered to re-express the relevant receptors showed normal levels of integrin activation (Fig. 1a). Basal levels of WOW-1-binding in VE-cadherin^{-/-} and PECAM-1^{-/-} endothelial cells were similar to those in reconstituted cells (data not shown). Sparse endothelial cells,

cultured at low cell-density and lacking cell-cell junctions, also showed no increase in WOW-1 binding after the onset of flow (data not shown).

We also assessed the events upstream of integrin activation in response to shear stress. VE-cadherin^{-/-} endothelial cells showed no significant increase in PI(3)K phosphorylation or activation of AKT, a PI(3)K-dependent event (Fig. 1b), consistent with a previous analysis of AKT in these cells⁶. Src activation was delayed but not substantially inhibited in VE-cadherin^{-/-} cells (Fig. 1c). However, PECAM-1^{-/-} endothelial cells showed no activation of PI(3)K, AKT or Src family kinases (Fig. 1a–c). Therefore, PECAM-1 is required for Src activation, while VE-cadherin is required for the transmission of that signal to PI(3)K.

Consistent with the idea that the integrin pathway mediates both cell alignment in the direction of flow and the transient induction of NF- κ B by flow³, neither VE-cadherin^{-/-} nor PECAM-1^{-/-} cell lines

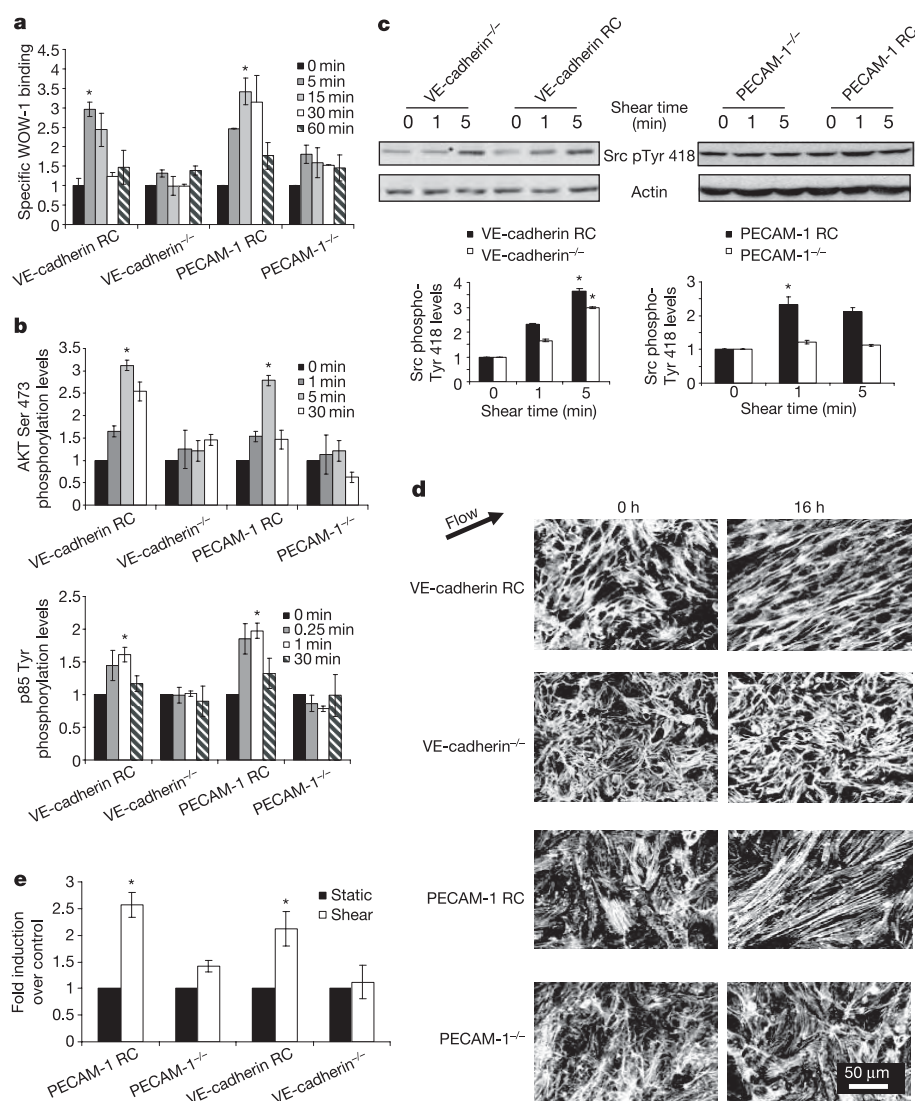


Figure 1 | Responses of PECAM-1^{-/-} and VE-cadherin^{-/-} cell lines to shear stress. **a**, PECAM-1^{-/-} and PECAM-1-reconstituted (PECAM-1 RC) cells, and VE-cadherin^{-/-} and VE-cadherin-reconstituted (VE-cadherin RC) cells, were subjected to shear stress for the times indicated and assayed for integrin activation by measuring WOW-1 binding. **b**, AKT activation was assayed by western blotting for phosphorylated AKT and total AKT. Values are means $\pm$ s.e.m.; $n = 3$. Phosphorylation of the PI(3)K p85 subunit was assayed by immunoprecipitation and western blotting for phosphorylated p85 as described in the Methods. **c**, Activity of Src family kinases was determined by western blotting for phospho-Tyr 418. Values are

means $\pm$ s.e.m.; $n = 3$. **d**, Cells were subjected to shear for 16 h then fixed and stained with TRITC-phalloidin to visualize actin filaments. The arrow indicates the direction of flow. **e**, The indicated endothelial cells were co-transfected with a vector containing the PDGF-A/SSRE regulating the expression of firefly luciferase, together with *Renilla* luciferase, under the control of a minimal promoter and subjected to shear for 60 min. Values represent firefly luciferase activity normalized to *Renilla* luciferase (mean $\pm$ s.e.m.; $n = 4$). Statistical significance as determined by Student's *t*-test is indicated by an asterisk ($P < 0.01$).

showed an alignment of actin filaments in the direction of flow (Fig. 1d), or an induction of luciferase under the control of an NF- κ B-dependent promoter¹⁴ (Fig. 1e). These data confirm the requirement for these adhesion receptors in the activation of the integrin pathway and downstream events by shear stress.

To test whether these receptors can directly transduce force, sparse endothelial cells were incubated with 4.5- μ m-diameter magnetic beads coated with antibodies to cell adhesion receptors, followed by the application of a magnetic field. Antibody-coated beads induced receptor binding, clustering and downstream signalling as assessed by their ability to recruit actin (Supplementary Fig. S2). The magnet exerts a maximum force on a 4.5- μ m-diameter bead of ~ 130 pN, which is $\sim 10\%$ of the drag force exerted by shear stress at 12 dyn cm^{-2} on a $30 \mu\text{m} \times 30 \mu\text{m}$ endothelial cell (~ 1.1 nN). Anti-PECAM-1-coated beads were bound to isolated bovine aortic endothelial cells expressing the GFP-AKT PH fusion protein as a sensor for PI(3)-lipids. Local activation of integrins was assessed by staining with the WOW-1 Fab fragment. The application of magnetic force for as little as 15 s triggered the activation of PI(3)K and integrins around the anti-PECAM-1-coated beads (Fig. 2a). The pretreatment of cells with PP2 or LY294002 abolished this response. Application of the same magnetic force to cells with bound anti-CD44- or anti-VE-cadherin-coated beads had little or no effect on activation levels (Fig. 2b). We also noted that when confluent endothelial cells were exposed to flow for 1 min, WOW-1 staining was higher near cell-cell junctions (Supplementary Fig. S3). However, at later timepoints WOW-1 stained the basal surface uniformly², indicating that either the active integrins or the activating signal could diffuse. We conclude that PECAM-1 is the direct transducer of mechanical force in this system.

When VE-cadherin^{-/-} endothelial cells were examined in this assay, the application of a magnetic force to bound anti-PECAM-1-coated beads failed to induce signalling (Fig. 2b), whereas reconstituted cells showed responses similar to bovine aortic endothelial cells

(data not shown). This result is surprising, because no ligand for VE-cadherin is present on the anti-PECAM-1-coated beads. To determine whether the role of VE-cadherin function in confluent monolayers under flow is also ligation-independent, VE-cadherin-reconstituted cells were mixed with a much larger number of VE-cadherin^{-/-} cells. Under these conditions, VE-cadherin cannot participate in homophilic adhesion. After 16 h of flow, isolated VE-cadherin-positive cells aligned normally, whereas surrounding VE-cadherin^{-/-} endothelial cells did not (Fig. 2c, d). Additionally, the treatment of cells in a confluent monolayer with a VE-cadherin-blocking antibody (BV9), which blocks homophilic binding and causes the dispersion of VE-cadherin from junctions to the rest of the cell surface¹⁵, induced the redistribution of VE-cadherin from intercellular junctions, but had no effect on flow-stimulated WOW-1 binding (data not shown). Thus, the involvement of VE-cadherin in shear-stress-dependent signalling is independent of cell-cell adhesion.

VEGFR2 shows rapid, Src-dependent, ligand-independent trans-activation in response to flow^{16,17}. Staining of cells with an antibody against VEGFR2 phosphotyrosine 1054, which is phosphorylated upon receptor activation, revealed the activation of VEGFR2 within 15 s upon the initiation of flow and localization of a fraction of the activated VEGFR2 to cell-cell junctions (Supplementary Fig. S4a). This signal persisted for at least 5 min and, consistent with previous results¹⁷, was Src-dependent (data not shown). Furthermore, flow-induced activation of VEGFR2 was absent in both PECAM-1^{-/-} and VE-cadherin^{-/-} endothelial cells at all times between 15 s and 5 min (Supplementary Fig. S4a; data not shown); reconstituted cells responded similarly to bovine aortic endothelial cells (data not shown). Therefore, VEGFR2 activation by shear stress is also downstream of the junctional receptors. We also noticed that β -catenin may translocate to the nucleus after the onset of flow in PECAM-1^{-/-} and VE-cadherin^{-/-} cells, but this observation was not investigated further. A selective inhibitor of VEGFR2 kinase activity (VTI)

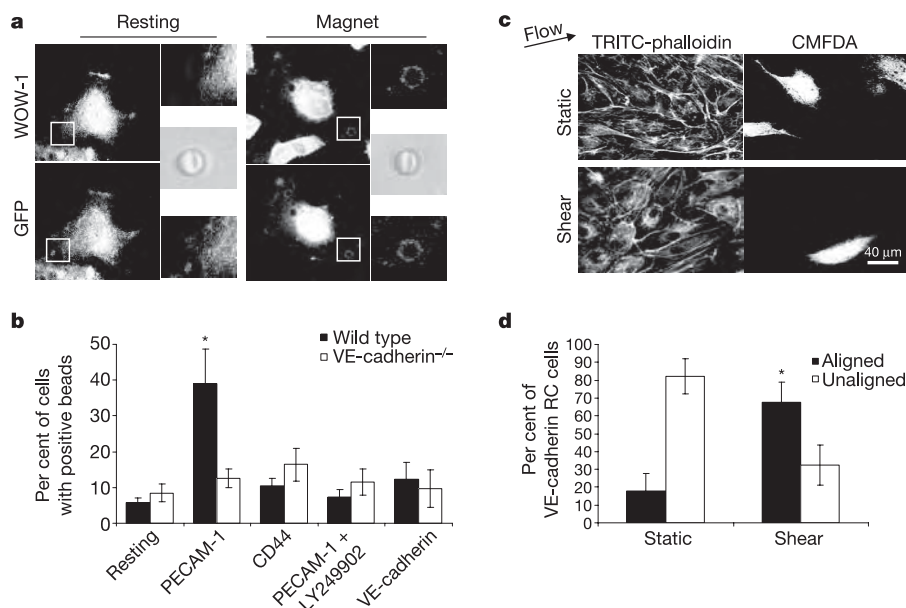


Figure 2 | Direct mechanotransduction and adhesion-independent role for VE-cadherin. **a, b**, Bovine aortic endothelial cells at low density and expressing the GFP-AKT PH fusion protein were incubated with antibody-coated magnetic beads for 30 min, in some cases after pretreatment with LY249902 to inhibit PI(3)K activity. Magnetic force was applied parallel to the coverslip for 15 s. **a**, Cells were fixed, stained with WOW-1 and examined for fluorescence. Insets show areas within the same field of view at a higher magnification. **b**, Antibody-coated beads were scored for a ring of WOW-1 or GFP-AKT PH fluorescence above background. Beads coated with anti-

VE-cadherin antibody and VE-cadherin^{-/-} cells were also analysed (>300 cells per condition). **c**, VE-cadherin-reconstituted cells loaded with CellTracker Green CMFDA dye and unlabelled VE-cadherin^{-/-} cells were mixed at a ratio of 1:100. Cells were sheared for 16 h, fixed and then stained with TRITC-phalloidin. Green cells were scored for alignment in the direction of flow (indicated by the arrow). **d**, Quantification of the alignment of VE-cadherin-reconstituted cells (200 cells counted). For experiments **a–c**, $n = 4$; for **d**, $n = 3$; values are mean $\pm$ s.e.m.; asterisk, $P < 0.01$.

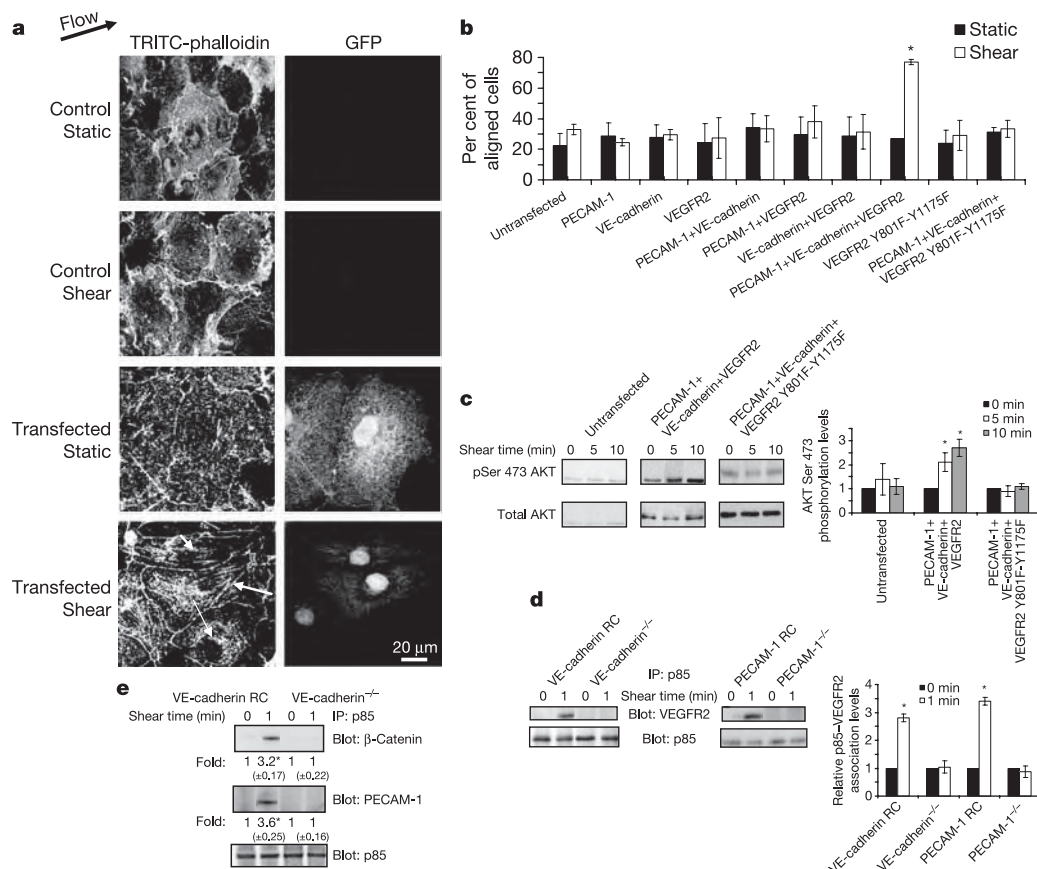


Figure 3 | PECAM-1, VE-cadherin and VEGFR2 form a mechanosensory complex. **a**, COS-7 cells transfected with constructs encoding PECAM-1, VE-cadherin and VEGFR2, together with GFP, were sheared for 16 h, fixed and stained with TRITC-phalloidin. Direction of flow is indicated by the arrow. White arrows show aligned stress fibres. **b**, Cells transfected with various combinations of receptor constructs, including the VEGFR2 Y801F-Y1175F double mutant, were analysed as in panel **a** and quantified. Values are means $\pm$ s.e.m.; $n = 2$; >100 cells were scored per condition. **c**, COS-7 cells were transiently transfected with PECAM-1, VE-cadherin and wild-type or mutant VEGFR2, and subjected to shear stress. Activation of

AKT was assayed by western blotting for phosphorylated AKT and total AKT. Values are means $\pm$ s.e.m.; $n = 3$. **d**, Lysates from VE-cadherin^{-/-} and VE-cadherin-reconstituted cells with or without shear were immunoprecipitated (IP) with an anti-PI(3)K p85 subunit antibody. VEGFR2 and p85 levels were assessed by western blotting. Values are means $\pm$ s.e.m.; $n = 3$. **e**, VE-cadherin^{-/-} and VE-cadherin-reconstituted cells were sheared, lysed and PI(3)K-immunoprecipitated. The fold induction of the indicated proteins was analysed by western blotting. Values are means $\pm$ s.e.m.; $n = 3$. For all experiments, an asterisk indicates $P < 0.01$.

blocked activation of PI(3)K in response to shear stress (Supplementary Fig. S4b). Flow-induced VEGFR2 activation is therefore required for PI(3)K activation.

The alignment of cells in the direction of flow is specific to endothelial cells¹⁸. To test whether PECAM-1, VE-cadherin and VEGFR2 are sufficient to transduce shear stress, COS-7 African green monkey cells transfected with various combinations of plasmids encoding PECAM-1, VE-cadherin and VEGFR2 were subjected to flow. Transfected cells expressed all three receptors (Supplementary Fig. S5a, b). Untransfected COS-7 cells did not align after 16 h of shear (Fig. 3a, b), nor did they activate AKT after shorter periods of shear exposure (Fig. 3c). In contrast, cells co-expressing all three proteins activated AKT and aligned in the direction of flow (Fig. 3a–c). None of the single- or double-transfectants showed any response in these assays. Thus, PECAM-1, VE-cadherin and VEGFR2 represent the essential endothelial-specific components needed for the alignment of cells by fluid flow.

VEGFR2 binds PI(3)K directly through phosphorylation sites at Tyr 801 and Tyr 1175 on the receptor, leading to the phosphorylation and activation of PI(3)K¹⁹. To test whether VEGFR2 interacts directly with PI(3)K in shear stress signalling, a Y801F-Y1175F double mutant was examined in COS-7 cells. This mutant receptor failed to induce AKT activation or cell alignment (Fig. 3b, c), suggesting that PI(3)K interacts directly with activated VEGFR2. To further

explore intermolecular associations, PI(3)K immunoprecipitates were analysed. VEGFR2 associated with PI(3)K only in cells that expressed VE-cadherin (Fig. 3d). Shear stress also induced association of PECAM-1 and β -catenin with the PI(3)K p85 subunit in a VE-cadherin-dependent manner (Fig. 3e). The association between VE-cadherin and VEGFR2 is thought to be indirect, mediated by β -catenin^{20,21}. β -catenin^{-/-} endothelial cells failed to show increased WOW-1 binding under flow (Supplementary Fig. S6). These results indicate that VE-cadherin and its binding partner β -catenin are required for the formation of signalling complexes that correlate with PI(3)K activation. Thus, VE-cadherin appears to function as an adaptor protein within this complex.

Elevated NF- κ B activity and expression of NF- κ B-dependent genes, including adhesion molecules such as ICAM-1, VCAM-1 and E-selectin, occurs in regions of disturbed flow before other indications of atherosclerosis and is believed to contribute to recruitment of monocytes to the nascent plaque^{22,23}. As NF- κ B activation and ICAM-1 expression depend on the integrin activation pathway *in vitro* (ref. 3), we examined NF- κ B and ICAM-1 at sites of disturbed flow *in vivo* in mice. Wild-type aortas had strong staining for ICAM-1 and nuclear translocation of NF- κ B at branch points (Fig. 4), consistent with published data²²; staining was undetectable in nearby areas where flow is laminar. In contrast, areas near branch points in the PECAM-1^{-/-} aortas showed no detectable activation of

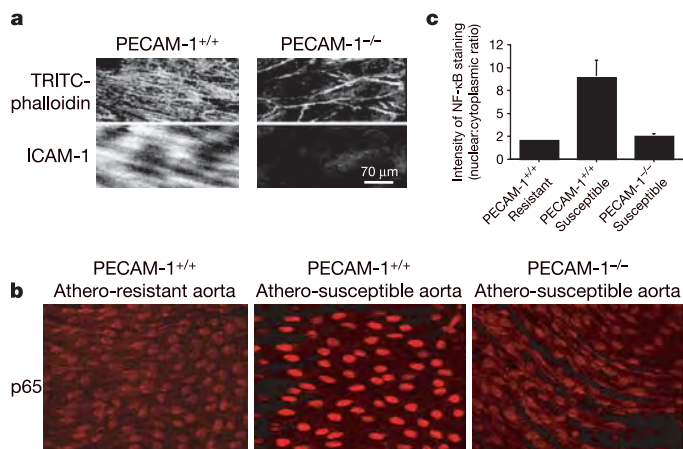


Figure 4 | Responses in PECAM-1^{-/-} mice. **a, b**, Fixed aortas from wild-type and PECAM-1^{-/-} mice were stained for F-actin (using TRITC-phalloidin) and ICAM-1 (**a**), and the p65 subunit of NF-κB (**b**). Images in panel **b** show an athero-resistant region or an athero-susceptible region of the aorta. **c**, NF-κB staining was quantified by determining nuclear and cytoplasmic intensity. Values are mean ± s.e.m.; *n* = 3; 30–60 cells analysed for each area per aorta.

NF-κB or expression of ICAM-1. Additionally, wild-type aortas showed extensive actin stress fibres across the cell, whereas PECAM-1^{-/-} aortas had most of the actin in a circumferential ring (Fig. 4). Thus, flow pathways that are dependent on integrin activation are impaired in PECAM-1^{-/-} mice.

Fluid shear stress triggers the conformational activation of integrins, which mediates both the alignment of endothelial cells in laminar shear and the activation of NF-κB in response to changes in shear^{2–4}. We have identified a mechanosensory complex comprised of PECAM-1 (which activates Src), VE-cadherin (which functions as an adaptor) and VEGFR2 (which activates PI(3)K). PI(3)K most probably leads to integrin activation through a conserved pathway shared with other cell types¹⁰. These events occurred within 15 s, consistent with a direct event. The PECAM-1 cytoplasmic domain can bind Src directly²⁴, suggesting that changes in this interaction in response to force may mediate mechanotransduction. These proteins represent the essential endothelial-cell-specific components for this pathway. Finally, PECAM-1^{-/-} mice show defects in both F-actin organization and activation of the NF-κB pathway, demonstrating its relevance *in vivo*.

Subconfluent endothelial cells still migrate selectively in the direction of flow²⁵ and non-endothelial cell types activate a subset of known flow-responsive pathways^{26,27}, so other mechanisms for shear stress sensing must exist. However, our data define a mechanosensing pathway that is critical for an important subset of known responses to flow. We note that polymorphisms in the PECAM-1 gene are associated with both decreased and increased incidences of artery disease²⁸. Therefore, investigating the effects of these mutations on PECAM-1 function in shear stress signalling may be worthwhile.

METHODS

Inhibitors, antibodies and dyes. VEGFR tyrosine kinase inhibitor (VTI, 4-[(4'-chloro-2'-fluoro)phenylamino]-6,7-dimethoxyquinazoline), PP2 (4-amino-5-(4-chlorophenyl)-7-(*t*-butyl)pyrazolo[3,4-d]pyrimidine), SU6656, wortmannin and LY294002 were from Calbiochem. Anti-phosphotyrosine (4G10) and anti-PI(3)K p85 antibodies were from Upstate Biotechnology. The BV9 monoclonal antibody to VE-cadherin¹⁵ and a polyclonal anti-VE-cadherin antibody were from BD Biosciences, the anti-β-catenin antibody was from Sigma, the monoclonal antibody to PECAM-1 was a gift from P. J. Newman. The anti-phospho-Src[pY 418], anti-phospho-Src[pY 527], anti-phospho-AKT[pS 473] and anti-phospho-VEGFR2[pY 1054] antibodies were from Biosource International. The anti-ICAM-1 antibody was from Zymed Laboratories, the 5D2-27

rat anti-mouse-CD44 antibody was from the Developmental Studies Hybridoma Bank, the antibody against the p65 subunit of NF-κB was from Transduction Laboratories, tetramethylrhodamine isothiocyanate (TRITC)- and fluorescein isothiocyanate-labelled goat anti-rabbit antibodies were from Jackson ImmunoResearch. CellTracker Green CMFDA (5-chloromethylfluorescein diacetate) dye was from Molecular Probes.

Cell culture, shear stress assays and transfections. Bovine aortic endothelial cells were cultured as previously described². VE-cadherin^{-/-} cells and cells reconstituted with human VE-cadherin were prepared as described²⁰. PECAM-1^{-/-} cells and cells reconstituted with full-length PECAM-1 were prepared as described^{29,30}. Levels of PECAM-1 and VE-cadherin in reconstituted cells are similar to wild-type levels^{20,29,30}. Bovine aortic endothelial cells were subjected to shear stress at 12 dyn cm⁻² in a parallel-plate flow chamber^{2–4}. Transfections were performed as described². The PECAM-1 complementary DNA in the pcDNA3 vector was a gift from P.J. Newman.

Alignment experiments. Confluent cells were subjected to flow for 16 h. In mixing experiments, reconstituted VE-cadherin endothelial cells were labelled with CellTracker Green CMFDA according to the manufacturer's instructions, mixed with unlabelled VE-cadherin^{-/-} cells at a ratio of 100:1 and subjected to shear stress for 16 h, then fixed and stained with TRITC-phalloidin.

Fluorescence microscopy. Staining of cells was carried out as described². Images of fixed cells were acquired using a 1024 confocal microscope (BioRad).

Luciferase activity assays. Luciferase activity assays were performed as described³ using a vector (1.0 μg) containing the PDGF-A-chain shear stress response element (PDGF-A/SSRE) regulating the expression of firefly luciferase, together with *Renilla* luciferase, under the control of a minimal promoter¹⁴. Cell cultures were electroporated by using the Electro cell manipulator (EC100, Fisher Scientific).

Immunoprecipitations. Cells were washed in ice-cold PBS containing Ca²⁺ and Mg²⁺ and collected in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM Na₃VO₄, 1% Triton X-100, 0.1% SDS, 10 mM NaF, 1 mM aprotinin, 1 mM phenylmethylsulphonyl fluoride, 1 μg ml⁻¹ leupeptin, 1 mM sodium pyrophosphate, 1 mM β-glycerophosphate). Samples were precleared with protein A or G sepharose (Pharmacia) for 1 h at 4°C. Supernatants were then incubated with protein A or G sepharose previously conjugated with the appropriate rabbit or mouse antibody for 1 h at 4°C under continuous mixing. Samples were washed three times in lysis buffer, and immune complexes were eluted in SDS sample buffer.

Direct force application. Magnetic beads (4.5 μm in diameter) conjugated with goat-anti-mouse IgG (Dynabeads; Dynal) were treated with 30 μg ml⁻¹ anti-PECAM-1, anti-CD44 or anti-VE-cadherin antibodies for 1 h and washed three times with PBS. Beads (2 × 10⁷) were added for 30 min to sparsely plated cells that had been transiently transfected with 0.5 μg of the construct encoding the GFP-AKT PH fusion protein. A magnet was passed over the cells for 15 s. Cells were fixed in formaldehyde and stained with the WOW-1 antibody Fab fragment and TRITC-anti-mouse IgG. In control experiments, cells were stained with rhodamine-phalloidin. The apparatus used to apply a high magnetic field gradient to the beads was comprised of eight cylindrically shaped NdFeB permanent magnets of 0.5 inches in diameter and 0.5 inches in length (part number 0013; Forcefield; <http://www.wondermagnet.com/>). Pairs of magnets were combined to make four units 0.5 inches wide and 1 inch long. These units were mounted in a lucite holder to form two columns so that a glass slide covered with cells could be passed through a gap. The magnet poles were configured so that the two magnets above the slide had their poles attracting each other, and the magnets located directly below the slide had their pair of poles oriented so as to produce repulsion between the top and the bottom set of magnets.

Staining aortas. Wild-type and PECAM-1^{-/-} C57BL/6 female mice were deeply anaesthetized with a mixture of ketamine and xylazine, and perfused with 20 ml of 0.9% NaCl into the ascending aorta for 1 min, followed by 100 ml of 4% paraformaldehyde for 5 min. The whole aorta was then removed and continually fixed within 4% paraformaldehyde for 4 h. The fixed aortas were kept in PBS at 4°C until staining.

Received 19 April; accepted 15 June 2005.

1. Davies, P. F. Overview: Temporal and spatial relationships in shear stress-mediated endothelial signalling. *J. Vasc. Res.* **34**, 208–211 (1997).
2. Tzima, E., del Pozo, M. A., Shattil, S. J., Chien, S. & Schwartz, M. A. Activation of integrins in endothelial cells by fluid shear stress mediates Rho-dependent cytoskeletal alignment. *EMBO J.* **20**, 4639–4647 (2001).
3. Tzima, E. *et al.* Activation of Rac1 by shear stress in endothelial cells mediates both cytoskeletal reorganization and effects on gene expression. *EMBO J.* **21**, 6791–6800 (2002).
4. Tzima, E., Kiosses, W. B., del Pozo, M. A. & Schwartz, M. A. Localized cdc42

- activation, detected using a novel assay, mediates microtubule organizing center positioning in endothelial cells in response to fluid shear stress. *J. Biol. Chem.* **278**, 31020–31023 (2003).
5. Newman, P. J. & Newman, D. K. Signal transduction pathways mediated by PECAM-1: new roles for an old molecule in platelet and vascular cell biology. *Arterioscler. Thromb. Vasc. Biol.* **23**, 953–964 (2003).
 6. Shay-Salit, A. *et al.* VEGF receptor 2 and the adherens junction as a mechanical transducer in vascular endothelial cells. *Proc. Natl Acad. Sci. USA* **99**, 9462–9467 (2002).
 7. Osawa, M., Masuda, M., Kusano, K. I. & Fujiwara, K. Evidence for a role of platelet endothelial cell adhesion molecule-1 in endothelial cell mechanosignal transduction: is it a mechanoresponsive molecule? *J. Cell Biol.* **158**, 773–785 (2002).
 8. Mohan, S., Mohan, N. & Sprague, E. A. Differential activation of NF- κ B in human aortic endothelial cells conditioned to specific flow environments. *Am. J. Physiol.* **273**, C572–C578 (1997).
 9. Monaco, C. & Paleolog, E. Nuclear factor κ B: A potential therapeutic target in atherosclerosis and thrombosis. *Cardiovasc. Res.* **61**, 671–682 (2004).
 10. Hughes, P. E. & Pfaff, M. Integrin affinity modulation. *Trends Cell Biol.* **8**, 359–364 (1998).
 11. Pampori, N. *et al.* Mechanisms and consequences of affinity modulation of integrin $\alpha_v\beta_3$ detected with a novel patch-engineered monovalent ligand. *J. Biol. Chem.* **274**, 21609–21616 (1999).
 12. Okuda, M. *et al.* Shear stress stimulation of p130(cas) tyrosine phosphorylation requires calcium-dependent c-Src activation. *J. Biol. Chem.* **274**, 26803–26809 (1999).
 13. Jalali, S. *et al.* Shear stress activates p60src-Ras-MAPK signalling pathways in vascular endothelial cells. *Arterioscler. Thromb. Vasc. Biol.* **18**, 227–234 (1998).
 14. Khachigian, L. M. *et al.* Egr-1 is activated in endothelial cells exposed to fluid shear stress and interacts with a novel shear-stress-response element in the PDGF-A-chain promoter. *Arterioscler. Thromb. Vasc. Biol.* **17**, 2280–2286 (1997).
 15. Corada, M. *et al.* Monoclonal antibodies directed to different regions of vascular endothelial cadherin extracellular domain affect adhesion and clustering of the protein and modulate endothelial permeability. *Blood* **97**, 1679–1684 (2001).
 16. Chen, K. D. *et al.* Mechanotransduction in response to shear stress. Roles of receptor tyrosine kinases, integrins, and Shc. *J. Biol. Chem.* **274**, 18393–18400 (1999).
 17. Jin, Z. G. *et al.* Ligand-independent activation of vascular endothelial growth factor receptor 2 by fluid shear stress regulates activation of endothelial nitric oxide synthase. *Circ. Res.* **93**, 354–363 (2003).
 18. Levesque, M. J. & Nerem, R. M. The elongation and orientation of cultured endothelial cells in response to shear stress. *J. Biomech. Eng.* **107**, 341–347 (1985).
 19. Dayanir, V., Meyer, R. D., Lashkari, K. & Rahimi, N. Identification of tyrosine residues in vascular endothelial growth factor receptor-2/FLK-1 involved in activation of phosphatidylinositol 3-kinase and cell proliferation. *J. Biol. Chem.* **276**, 17686–17692 (2001).
 20. Lampugnani, M. G. *et al.* VE-cadherin regulates endothelial actin activating Rac and increasing membrane association of Tiam. *Mol. Biol. Cell* **13**, 1175–1189 (2002).
 21. Grazia Lampugnani, M. *et al.* Contact inhibition of VEGF-induced proliferation requires vascular endothelial cadherin, beta-catenin, and the phosphatase DEP-1/CD148. *J. Cell Biol.* **161**, 793–804 (2003).
 22. Nakashima, Y., Raines, E. W., Plump, A. S., Breslow, J. L. & Ross, R. Upregulation of VCAM-1 and ICAM-1 at atherosclerosis-prone sites on the endothelium in the ApoE-deficient mouse. *Arterioscler. Thromb. Vasc. Biol.* **18**, 842–851 (1998).
 23. Hajra, L. *et al.* The NF-kappa B signal transduction pathway in aortic endothelial cells is primed for activation in regions predisposed to atherosclerotic lesion formation. *Proc. Natl Acad. Sci. USA* **97**, 9052–9057 (2000).
 24. Lu, T. T., Barreuther, M., Davis, S. & Madri, J. A. Platelet endothelial cell adhesion molecule-1 is phosphorylatable by c-Src, binds Src-Src homology 2 domain, and exhibits immunoreceptor tyrosine-based activation motif-like properties. *J. Biol. Chem.* **272**, 14442–14446 (1997).
 25. Masuda, M. & Fujiwara, K. Morphological responses of single endothelial cells exposed to physiological levels of fluid shear stress. *Front. Med. Biol. Eng.* **5**, 79–87 (1993).
 26. Smalt, R., Mitchell, F. T., Howard, R. L. & Chambers, T. J. Mechanotransduction in bone cells: induction of nitric oxide and prostaglandin synthesis by fluid shear stress, but not by mechanical strain. *Adv. Exp. Med. Biol.* **433**, 311–314 (1997).
 27. van der Pauw, M. T. *et al.* Response of periodontal ligament fibroblasts and gingival fibroblasts to pulsating fluid flow: nitric oxide and prostaglandin E2 release and expression of tissue non-specific alkaline phosphatase activity. *J. Periodontal Res.* **35**, 335–343 (2000).
 28. Elayess, M. A. *et al.* A novel functional polymorphism in the PECAM-1 gene (53G > A) is associated with progression of atherosclerosis in the LOCAT and REGRESS studies. *Atherosclerosis* **168**, 131–138 (2003).
 29. Wong, C. W. *et al.* PECAM-1/CD31 trans-homophilic binding at the intercellular junctions is independent of its cytoplasmic domain; evidence for heterophilic interaction with integrin $\alpha_v\beta_3$ in *cis*. *Mol. Biol. Cell* **11**, 3109–3121 (2000).
 30. Graesser, D. *et al.* Altered vascular permeability and early onset of experimental autoimmune encephalomyelitis in PECAM-1-deficient mice. *J. Clin. Invest.* **109**, 383–392 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by the NIH. E.T. is an AHA Western States Fellow. H.D. is supported by the Department of Defense. We thank M. Ginsberg, D. Salomon, J. Quigley and R. Klemke for their help. We also thank N. Resnick for providing the PDGF-A/SSRE construct, E. Schaefer for providing the phospho-specific VEGFR2 antibodies and J. Downward for the GFP-AKT PH construct. We thank P. J. Newman, S. Chien and the Developmental Studies Hybridoma Bank for providing additional reagents. Discussions with A. W. Orr and J. S. Reader were also appreciated.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A.S. (maschwartz@virginia.edu).

LETTERS

An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation

Min Gyu Lee¹, Christopher Wynder¹, Neil Cooch¹ & Ramin Shiekhattar¹

We have previously described a multiprotein complex termed the BHC or BRAF–HDAC complex, which is required for the repression of neuronal-specific genes¹. We have shown that the BHC complex is recruited by a neuronal silencer, REST (RE1-silencing transcription factor), and mediates the repression of REST-responsive genes¹. BHC is a multiprotein complex consisting of two enzymatic activities: a histone deacetylase (HDAC1 or 2) and a recently described histone demethylase (BHC110, also known as LSD1 or AOF2)^{1–3}. Here we show that BHC110-containing complexes show a nearly fivefold increase in demethylation of histone H3 lysine 4 (H3K4) compared to recombinant BHC110. Furthermore, recombinant BHC110 is unable to demethylate H3K4 on nucleosomes, but BHC110-containing complexes readily demethylate nucleosomes. *In vitro* reconstitution of the BHC complex using recombinant subunits reveals an essential role for the REST corepressor CoREST, not only in stimulating demethylation on core histones but also promoting demethylation of nucleosomal substrates. We find that nucleosomal demethylation is the result of CoREST enhancing the association between BHC110 and nucleosomes. Depletion of CoREST in *in vivo* cell culture results in de-repression of REST-responsive gene expression and increased methylation of H3K4. Together, these results highlight an essential role for CoREST in demethylation of H3K4 both *in vitro* and *in vivo*.

Repression of neuronal-specific genes in non-neuronal cells requires the recruitment of the transcriptional corepressor complex BHC by the neuronal silencer protein REST¹. The BHC complex is composed of six subunits, including the histone deacetylase HDAC1/2, the SANT-domain-containing protein CoREST, the PHD-domain-containing protein BHC80, the HMG-domain-containing protein BRAF35, and the recently described histone demethylase BHC110 (also known as LSD1) (refs 1, 2), which is responsible for demethylating mono- and di-methyl histone H3 lysine 4 (H3K4). We have previously shown that in addition to the BHC complex, BHC110 and HDAC1/2 define a family of multiprotein complexes that share these two core enzymatic subunits³. To define the enzymatic activities of BHC110-containing complexes, we developed HEK-293-derived cell lines expressing Flag-tagged BHC110. We also generated stable cell lines expressing Flag-BHC110 containing a lysine-to-alanine (K661A) mutation. This conserved lysine residue has been shown to direct essential interactions with the flavin cofactor through water molecules in the monoamine oxidase crystal structure^{4,5} (Fig. 1a).

We isolated wild-type BHC110 and BHC110 (K661A) mutant complexes using Flag-affinity resin. Analysis of the affinity eluate by SDS-PAGE followed by silver staining and mass spectrometric analysis revealed a similar polypeptide composition for the two complexes (Fig. 1b). We next assessed the enzymatic activities of the two complexes in histone demethylation and deacetylation assays. At equal concentrations, both complexes showed similar

deacetylation of histone H3, but the K661A mutation of BHC110 completely abrogated histone demethylation activity (Fig. 1c). These results support a role for BHC110 as the histone demethylase enzymatic activity of BHC110-containing complexes.

We next compared the histone demethylase activity for recombinant BHC110 isolated from insect cells to that of BHC110-containing complexes, using core histones (Fig. 2a, b). This analysis revealed an approximately fivefold enhancement of H3K4 demethylation by BHC110-containing complexes compared with recombinant enzyme (Fig. 2b, c). Furthermore, BHC110-containing complexes demethylate nucleosomal H3K4, but recombinant BHC110 was completely inert towards nucleosomal substrates (Fig. 2d, e). These results show that proteins associated with BHC110 in the complex not only

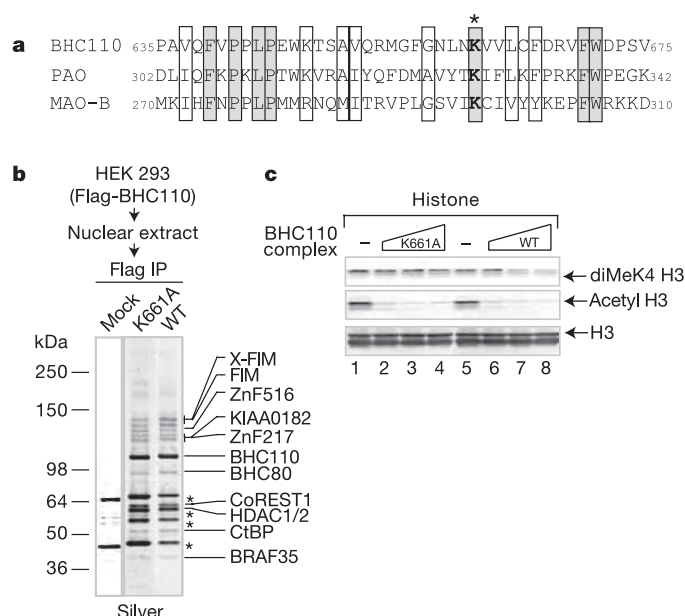


Figure 1 | Mutation in a conserved lysine residue abrogates demethylation activity of BHC110-containing complexes. **a**, Sequence alignment of BHC110 with polyamine oxidase (PAO) and monoamine oxidase B (MAO-B). Identical or similar residues are shaded in grey or white boxes, respectively, and the conserved lysine residues are marked by an asterisk. **b**, Analysis of wild-type (WT) and mutant (K661A) BHC110-containing complexes isolated from nuclear extract by silver staining (right panel). Left panel depicts silver staining of the Flag-affinity eluate from control untagged HEK-293 cells (mock). Asterisks denote nonspecific polypeptides. BHC110-associated polypeptides (shown on the right) have been previously described³. **c**, Comparison of demethylation and deacetylation activities in wild-type and mutant BHC110-containing complexes. diMeK4 H3, dimethyl-K4 H3; Acetyl H3, acetyl-K9/K14 H3.

¹The Wistar Institute, 3601 Spruce Street, Philadelphia, Pennsylvania 19104, USA.

promote the demethylation of core histones by BHC110, but also allow BHC110 to demethylate nucleosomal substrates.

To define the subunit(s) involved in modulating the activity of BHC110 for H3K4 demethylation, we reconstituted the previously defined BHC complex using recombinant subunits produced in insect cells (Fig. 3a). Addition of HDAC1 resulted in a small increase in the demethylation of core histones by BHC110 (compare lanes 3 and 6 in Fig. 3b; see also Supplementary Fig. 1a). Addition of CoREST not only enhanced H3K4 demethylation, but also yielded increased H3 deacetylation activity on core histone substrates (Fig. 3b). Addition of BHC80 and BRAF35 did not increase the enzymatic activity of BHC110 and HDAC1 above that observed with the addition of CoREST (Fig. 3b, lanes 11–18). Notably, inhibition of HDAC1 activity did not affect demethylation of core histones by BHC110 (Supplementary Fig. 1b).

None of the individual BHC components resulted in nucleosomal H3K4 demethylation (Fig. 3c), but reconstitution of BHC revealed an essential role for CoREST in mediating nucleosomal H3K4 demethylation (Fig. 3d). Although other subunits of the BHC complex (BHC80 and BRAF35) did not affect CoREST-induced nucleosomal demethylation, they increased the deacetylation of H3 by HDAC1 (Fig. 3d, compare lane 4 with lanes 5 and 6). These results highlight a necessary role for CoREST in mediating nucleosomal H3K4 demethylation. CoREST, BHC80 and BRAF35 also potentiate

the histone deacetylase activity of HDAC1 on nucleosomal substrates, reflecting additional regulatory roles for these proteins in the corepressor function of BHC.

To assess directly the role of CoREST in mediating nucleosomal H3K4 demethylation, we analysed the physical and functional association between BHC110 and CoREST (Fig. 4a). Recombinant BHC110 and CoREST were mixed, and this mixture was analysed by fractionation on a Superdex 200 gel filtration column (Fig. 4a). Analysis of gel filtration column fractions by SDS-PAGE, silver staining and western blot analysis revealed a physical association between CoREST and BHC110 (compare fraction 18 to fractions 24–26 in Fig. 4a; data not shown). Adding increasing concentrations of CoREST to recombinant BHC110 was sufficient to induce nucleosomal demethylation (Fig. 4b). Moreover, deletion mapping analysis of CoREST revealed that both SANT domains are required not only for association with BHC110 but also for CoREST-mediated nucleosomal demethylation (Supplementary Fig. 2). Together, these results point to a direct role for CoREST in mediating nucleosomal demethylation by BHC110, and suggest a redundant role for the two SANT domains in mediating BHC110 association.

To examine whether the effects of CoREST on nucleosome demethylation required oligonucleosomes, or whether such effects

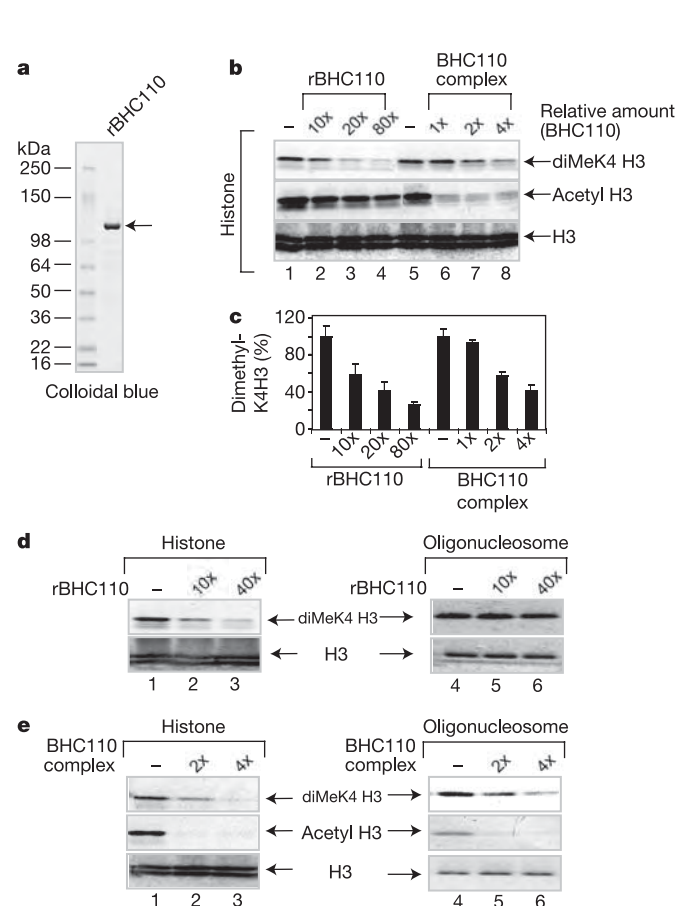


Figure 2 | Demethylation of nucleosomes by BHC110 complexes. **a**, Colloidal blue staining of recombinant BHC110 (rBHC110) isolated from Sf21 insect cells. **b**, Comparison of demethylation and deacetylation activities in recombinant BHC110 and BHC110-containing complex. **c**, Quantification of diMeK4 H3 levels are shown. Dimethyl-K4 levels in the absence of BHC110 (–) were set as 100%. Data are represented as mean $\pm$ s.d. of three experiments. **d**, Demethylation of histones and nucleosomes using recombinant BHC110. **e**, Demethylation of histones and nucleosomes using BHC110-containing complex. '1x' corresponds to 50 ng of recombinant BHC110 or BHC110 in the complex.

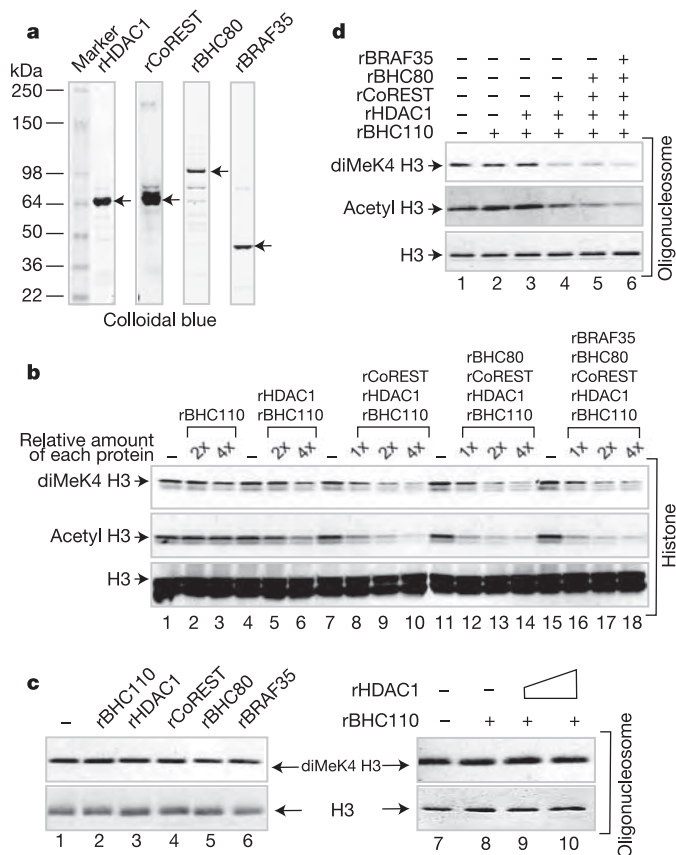


Figure 3 | CoREST promotes nucleosomal demethylation *in vitro*. **a**, Colloidal blue staining of recombinant (r)HDAC1, rCoREST, rBHC80 and rBRAF35 isolated from Sf21 insect cells. **b**, Demethylation of histones using different combination of recombinant proteins. '1x' corresponds to 150 ng rBHC110, 75 ng rHDAC1, 75 ng rCoREST, 60 ng rBHC80 or 60 ng rBRAF35. **c**, Demethylation of nucleosomes using individual recombinant proteins (left panel) and recombinant BHC110 plus HDAC1 (right panel). The following amounts of recombinant proteins were used: 800 ng rBHC110, 400–800 ng rHDAC1, 400 ng rCoREST, 300 ng rBHC80, 300 ng rBRAF35. **d**, Demethylation of nucleosomes using different combinations of recombinant proteins. The amounts of recombinant proteins used correspond to '4x' as described in panel b.

could be also obtained using mononucleosomes, we treated oligonucleosomes using micrococcal nuclease to generate mononucleosomes. Similar to the results obtained with oligonucleosomes, addition of HDAC1 did not affect demethylation of mononucleosomes, and CoREST was required to achieve demethylation of mononucleosomes by BHC110 (Fig. 4c). Analysis of the physical association between BHC110 and mononucleosomes showed that although BHC110 alone did not associate with nucleosomes, CoREST associated weakly with nucleosomes and promoted nucleosomal interaction in the presence of BHC110 (Fig. 4d). Together, these results identify a role for CoREST in mediating the interaction between the histone demethylase and nucleosome to promote H3K4 demethylation.

To determine whether CoREST has a similar role in demethylation *in vivo*, we used small interfering RNA to deplete CoREST (Fig. 4e). Reduction of CoREST levels led to decreased levels of CoREST and BHC110 at the synapsin promoter, and a concomitant increase in REST-responsive gene expression (Fig. 4e, f). Moreover, CoREST depletion resulted in a nearly fourfold increase in H3K4 methylation (Fig. 4f). Similar enhancements of REST-responsive genes were obtained by siRNA-mediated depletion of BHC110 (Supplementary Fig. 3). These results support a role for CoREST in modulating the

in vivo association of BHC110 with the promoter region of REST-responsive genes.

CoREST was first identified through its association with the neuronal silencer REST⁶. We and others have subsequently identified CoREST as a component of a multiprotein complex required for REST-responsive gene repression^{1,3,6,7–14}. Here we show a role for CoREST in mediating nucleosomal demethylation by BHC110. Although the experiments described here identify an essential role for CoREST in demethylation of dimethyl-K4, similar results were obtained for the role of CoREST in demethylation of monomethyl-K4 (data not shown). Moreover, our results suggest a role for CoREST in promoting histone deacetylation by the BHC complex (see Fig. 3d).

CoREST not only shares structural similarities (a pair of SANT motifs) with other corepressor proteins, such as the nuclear receptor corepressors SMRT and N-CoR^{15–18}, but also contains an ELM2 domain that might mediate its interaction with HDAC1 (ref. 19). Such SANT domains have been shown to promote histone deacetylation as well as mediating interactions with histone tails^{15,16,20}. Our analysis of CoREST SANT domains indicates a requirement for both SANT domains in mediating the association of CoREST with BHC110 and promoting nucleosomal demethylation activity. Future crystallographic analysis of CoREST–BHC110–nucleosome complexes might provide further insights into the functions of different CoREST domains.

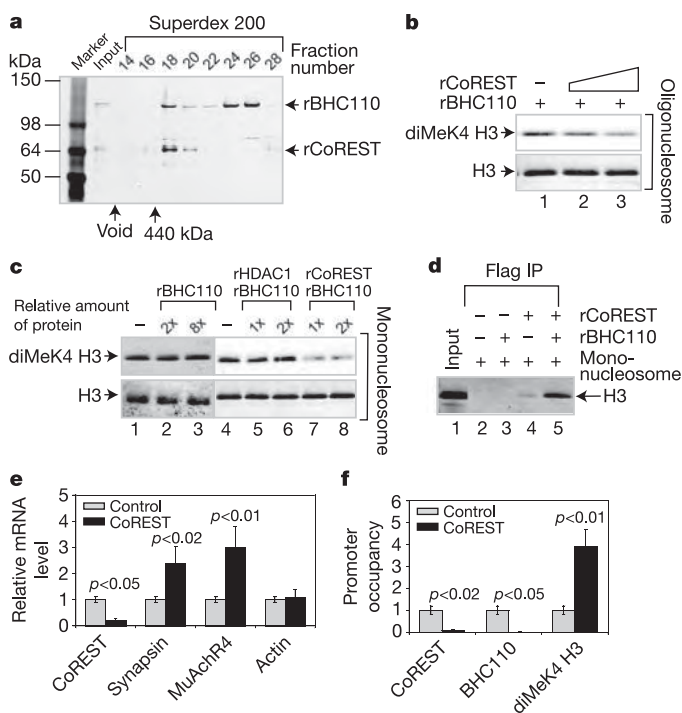


Figure 4 | CoREST mediates the association of BHC110 with nucleosomes.

a, Analysis of recombinant BHC110 and CoREST by gel filtration on a Superdex 200 column followed by silver staining. 'Void' represents the molecular weight cutoff for the column. **b**, Effect of recombinant CoREST (200–400 ng) on nucleosomal demethylation by recombinant BHC110 (800 ng). **c**, Demethylation assay on mononucleosomes using recombinant BHC110, HDAC1 and CoREST. "1 ×" corresponds to 400 ng BHC110 and 200 ng of CoREST and HDAC1. **d**, Western blot analysis of mononucleosomes co-immunoprecipitated with BHC110 and CoREST. Both recombinant proteins contain a Flag-epitope. **e**, mRNA levels of CoREST, synapsin, muscarinic acetylcholine receptor M4 (MuAChR4) and actin, measured by quantitative PCR with reverse transcription after treatment of HEK-293 cells with siRNA against CoREST ($n = 3$). siRNA against luciferase was used as control. **f**, Analysis of CoREST, BHC110 and dimethyl-K4 H3 occupancy of the synapsin promoter, by quantitative chromatin immunoprecipitation assay after treatment of HEK-293 cells with siRNA against CoREST ($n = 4$). Error bars represent s.e.m. for four experiments.

METHODS

Plasmids, histones, nucleosomes and other reagents. Wild-type human BHC110 and the K661A mutant were cloned into a pFLAG-CMV2 vector (Sigma). Baculoviral transfer vectors encoding Flag-BHC110, Flag-CoREST-(His)₆, Flag-BHC80 and Flag-BRAF35 were generated using pBlueBac4.5 plasmid (Invitrogen). Bacterial expression plasmids encoding Flag-CoREST or FLAG-CoREST with deleted ELM2 or SANT domains (Flag-ΔELM2, Flag-ΔSANT2, Flag-ΔSANT1 and Flag-ΔSANT1 + 2) were generated using pET28a (Novagen). Bulk histones were purchased from Sigma (H9250). Nucleosomes were purified from a HeLa nuclear pellet as previously described²¹. Micrococcal nuclease was used to digest nucleosomes according to the manufacturer's instructions (Roche). Anti-dimethyl K4H3 (12-460), anti-monomethyl K4H3 (07-436) and anti-K9/K14 acetyl (06-599) antibodies were purchased from Upstate. The anti-H3 antibody (ab1791) was from Abcam Ltd. Anti-Flag antibody (F3165) was from Sigma.

Affinity purification. Baculoviral recombinant proteins (Flag-BHC110, Flag-HDAC1, Flag-CoREST-(His)₆, Flag-BHC80 and Flag-BRAF 35) were purified from Sf21 insect cells infected by recombinant viruses using anti-Flag M2 affinity resin (Sigma) as previously described²². Bacterial recombinant proteins (Flag-CoREST, Flag-ΔELM2, Flag-ΔSANT2, Flag-ΔSANT1 and Flag-ΔSANT1 + 2) were similarly purified from BL21 cells. The wild-type or mutant BHC110-containing complex was purified from 150–200 mg nuclear extract isolated from the stable cell lines using anti-Flag M2 affinity resin as previously described²³. BHC110-associated proteins were identified by liquid chromatography–tandem mass spectroscopy. The amount of BHC110 in complexes was determined by silver staining, and recombinant proteins were determined by colloidal blue staining compared with known amounts of BSA.

Demethylation and deacetylation assay. Bulk histones (4 μg) were incubated with the indicated amounts of recombinant proteins or BHC110 complexes in histone demethylase (HDM) assay buffer A (50 mM Tris pH 8.5, 50 mM KCl, 5 mM MgCl₂, 5% glycerol, 0.2 mM phenylmethylsulphonyl fluoride and 1 mM dithiothreitol) in a final volume of 10 μl for 12–16 h at 37 °C. For nucleosome (0.3 μg) or mononucleosome (0.3 μg), HDM buffer A containing 0.1% NP40 was used. The reaction mixture was analysed by SDS–PAGE and western blotting. Antibodies against di- (or mono-) methyl K4H3 and acetyl-K9/K14H3 were used to detect methylation and acetylation levels, respectively.

In vitro interaction assays. To test the *in vitro* interaction between BHC110 and CoREST, recombinant BHC110 (10 μg) and CoREST (5 μg) were mixed and incubated for 1 h at 4 °C, fractionated by Superdex 200 gel filtration column in a buffer containing 20 mM Tris-HCl pH 7.9, 500 mM KCl, 10% glycerol, 0.2 mM EDTA, 1 mM dithiothreitol, 0.1% Nonidet P40 and 0.2 mM phenylmethylsulphonyl fluoride, and then analysed by silver staining. For co-immunoprecipitation of mononucleosomes with BHC110 and CoREST, nucleosomes (1.5 μg) were digested with micrococcal nuclease and incubated with recombinant BHC110 (1 μg), CoREST (500 ng) or both proteins in HDM buffer A containing

0.1% NP40 for 1 h at 4 °C. Anti-Flag M2 affinity resin was then added. The beads were washed extensively with HDM buffer A containing 0.1% NP40 and eluted with wash buffer containing 0.5 µg ml⁻¹ Flag peptide. Eluates were analysed by SDS-PAGE and western blotting.

RNA interference, quantitative RT-PCR and chromatin immunoprecipitation. RNA interference was performed according to the manufacturer's instructions using lipofectamine 2000 (Invitrogen). Short interfering (si)RNAs were purchased from Dharmacon's siGENOME collection. The sequences of two CoREST siRNAs (sense strand) were 5'-GGAAUUGGUUCAGUCAAUUU-3' and 5'-CGACGCGCUUCAACAUAGUU-3'. The sequence of the luciferase siRNA (control) sense strand was 5'-AACGUACGCGAAUACUUCGA-3'. RNA was prepared using the Qiagen RNeasy kit and reverse-transcribed using the Invitrogen First Strand Synthesis kit. Quantitative polymerase chain reaction (PCR) was carried out using Opticon2 (MJ Research) with DyNAmo HS SYBR Green qPCR kit (Finnzymes). Each sample was analysed in triplicate for both GAPDH and either synapsin or CoREST (with primers 5'-GGCACTCGGCATGCTTCT-3', 5'-CTAACTCACCTGCCACCT-3'), and quantified using opticon software. Synapsin or CoREST messenger RNA levels were normalized to GAPDH levels. Relative mRNA levels represent fold increase over control. Data in Fig. 4e, f are presented as mean ± s.e.m. Chromatin immunoprecipitation assays were performed as described³. Promoter occupancy levels were measured with the DyNAmo HS SYBR Green qPCR kit using Opticon2, and are expressed as fold increase over control.

Received 11 April; accepted 8 July 2005.

Published online 3 August 2005.

- Hakimi, M. A. *et al.* A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc. Natl Acad. Sci. USA* **99**, 7420–7425 (2002).
- Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).
- Hakimi, M. A., Dong, Y., Lane, W. S., Speicher, D. W. & Shiekhattar, R. A candidate X-linked mental retardation gene is a component of a new family of histone deacetylase-containing complexes. *J. Biol. Chem.* **278**, 7234–7239 (2003).
- Binda, C. *et al.* A 30-angstrom-long U-shaped catalytic tunnel in the crystal structure of polyamine oxidase. *Struct. Fold. Des.* **7**, 265–276 (1999).
- Binda, C. *et al.* Insights into the mode of inhibition of human mitochondrial monoamine oxidase B from high-resolution crystal structures. *Proc. Natl Acad. Sci. USA* **100**, 9750–9755 (2003).
- Andres, M. E. *et al.* CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proc. Natl Acad. Sci. USA* **96**, 9873–9878 (1999).
- Humphrey, G. W. *et al.* Stable histone deacetylase complexes distinguished by the presence of SANT domain proteins CoREST/kiaa0071 and Mta-L1. *J. Biol. Chem.* **276**, 6817–6824 (2001).
- Battaglioli, E. *et al.* REST repression of neuronal genes requires components of the hSWI.SNF complex. *J. Biol. Chem.* **277**, 41038–41045 (2002).
- Ballas, N. *et al.* Regulation of neuronal traits by a novel transcriptional complex. *Neuron* **31**, 353–365 (2001).
- Grimes, J. A. *et al.* The co-repressor mSin3A is a functional component of the REST-CoREST repressor complex. *J. Biol. Chem.* **275**, 9461–9467 (2000).
- Yamagoe, S. *et al.* Interaction of histone acetylases and deacetylases *in vivo*. *Mol. Cell. Biol.* **23**, 1025–1033 (2003).
- You, A., Tong, J. K., Grozinger, C. M. & Schreiber, S. L. CoREST is an integral component of the CoREST-human histone deacetylase complex. *Proc. Natl Acad. Sci. USA* **98**, 1454–1458 (2001).
- Lunyak, V. V. *et al.* Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science* **298**, 1747–1752 (2002).
- Dallman, J. E., Allopenna, J., Bassett, A., Travers, A. & Mandel, G. A conserved role but different partners for the transcriptional corepressor CoREST in fly and mammalian nervous system formation. *J. Neurosci.* **24**, 7186–7193 (2004).
- Yu, J., Li, Y., Ishizuka, T., Guenther, M. G. & Lazar, M. A. A SANT motif in the SMRT corepressor interprets the histone code and promotes histone deacetylation. *EMBO J.* **22**, 3403–3410 (2003).
- Boyer, L. A. *et al.* Essential role for the SANT domain in the functioning of multiple chromatin remodeling enzymes. *Mol. Cell* **10**, 935–942 (2002).
- Aasland, R., Stewart, A. F. & Gibson, T. The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIB. *Trends Biochem. Sci.* **21**, 87–88 (1996).
- Guenther, M. G., Barak, O. & Lazar, M. A. The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. *Mol. Cell. Biol.* **21**, 6091–6101 (2001).
- Ding, Z., Gillespie, L. L. & Paterno, G. D. Human MI-ER1 alpha and beta function as transcriptional repressors by recruitment of histone deacetylase 1 to their conserved ELM2 domain. *Mol. Cell. Biol.* **23**, 250–258 (2003).
- Sterner, D. E., Wang, X., Bloom, M. H., Simon, G. M. & Berger, S. L. The SANT domain of Ada2 is required for normal acetylation of histones by the yeast SAGA complex. *J. Biol. Chem.* **277**, 8178–8186 (2002).
- Barak, O. *et al.* Isolation of human NURF: a regulator of *Engrailed* gene expression. *EMBO J.* **22**, 6089–6100 (2003).
- Dong, Y. *et al.* Regulation of BRCC, a holoenzyme complex containing BRCA1 and BRCA2, by a signalosome-like subunit and its role in DNA repair. *Mol. Cell* **12**, 1087–1099 (2003).
- Gregory, R. I. *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235–240 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank D. Bochar and M. A. Hakimi for generating baculoviruses encoding BHC subunits. We are grateful to R. Gregory for comments on the manuscript. R.S. was supported by a grant from the American Cancer Society (ACS).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.S. (shiekhattar@wistar.org).

LETTERS

LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription

Eric Metzger¹, Melanie Wissmann^{1*}, Na Yin^{1*}, Judith M. Müller¹, Robert Schneider², Antoine H. F. M. Peters³, Thomas Günther¹, Reinhard Büttner⁴ & Roland Schüle¹

Gene regulation in eukaryotes requires the coordinate interaction of chromatin-modulating proteins with specific transcription factors such as the androgen receptor¹. Gene activation and repression is specifically regulated by histone methylation status at distinct lysine residues². Here we show that lysine-specific demethylase 1 (LSD1; also known as BHC110)³ co-localizes with the androgen receptor in normal human prostate and prostate tumour. LSD1 interacts with androgen receptor *in vitro* and *in vivo*, and stimulates androgen-receptor-dependent transcription. Conversely, knockdown of LSD1 protein levels abrogates androgen-induced transcriptional activation and cell proliferation. Chromatin immunoprecipitation analyses demonstrate that androgen receptor and LSD1 form chromatin-associated complexes in a ligand-dependent manner. LSD1 relieves repressive histone marks by demethylation of histone H3 at lysine 9 (H3-K9), thereby leading to de-repression of androgen receptor target genes. Furthermore, we identify pargyline as an inhibitor of LSD1. Pargyline blocks demethylation of H3-K9 by LSD1 and consequently androgen-receptor-dependent transcription. Thus, modulation of LSD1 activity offers a new strategy to regulate androgen receptor functions. Here, we link demethylation of a repressive histone mark with androgen-receptor-dependent gene activation, thus providing a mechanism by which demethylases control specific gene expression.

Transcriptional regulation by nuclear receptors such as androgen receptor (AR) involves interaction with multiple factors that act in both a sequential and combinatorial manner to reorganize chromatin¹. Central to this dynamic organization is the modification of core histones. The amino-terminal tails of histones are subject to various covalent modifications such as acetylation, phosphorylation, ubiquitination and methylation by specific chromatin-modifying enzymes². Histone methylation at specific lysine residues is linked to both transcriptional repression and activation².

We isolated LSD1 (ref. 3) by searching for new AR-interacting proteins. Endogenous LSD1 and AR associate specifically *in vivo* in androgen-sensitive tissues such as testis (Fig. 1a). LSD1 contains a centrally located SWIRM domain, which functions as a putative protein–protein interaction motif, and a carboxy-terminal amine oxidase domain that harbours the demethylase activity³ (Fig. 1b). As shown in glutathione S-transferase (GST) pull-down analyses, full-length LSD1, as well as the SWIRM domain (LSD1 175–246) and the amine oxidase domain (LSD1 247–852) alone, associate with the N terminus (NTD), the DNA-binding domain (DBD) and the ligand-binding domain (LBD) of AR (Fig. 1b). In contrast, the N terminus of LSD1 (LSD1 1–174) does not interact with AR. Furthermore, neither LSD1 nor the LSD1 mutants associate with GST, GST–Nix1,

GST–ROR β or GST–ER β -NTD, thus demonstrating specificity of interaction with AR.

To examine the expression pattern of LSD1, we performed northern blot analyses. LSD1 messenger RNA is ubiquitously expressed in human and murine fetal and adult tissue (Fig. 2a and data not shown) as a transcript of 3.3 kilobases (Supplementary Fig. S2a). To investigate LSD1 localization in normal prostate and prostate tumours, we used immunohistochemical analyses of 100 prostate cancer biopsies on tissue microarrays. As shown exemplarily in Fig. 2b, LSD1 is detected in the epithelium of normal prostate and in tumour cells. Importantly, these cells also express AR (Fig. 2b), showing that LSD1 and AR co-localize. The nuclear co-localization of LSD1 and AR was verified further in human LNCaP prostate tumour cells (Supplementary Fig. S2b). Taken together, our data

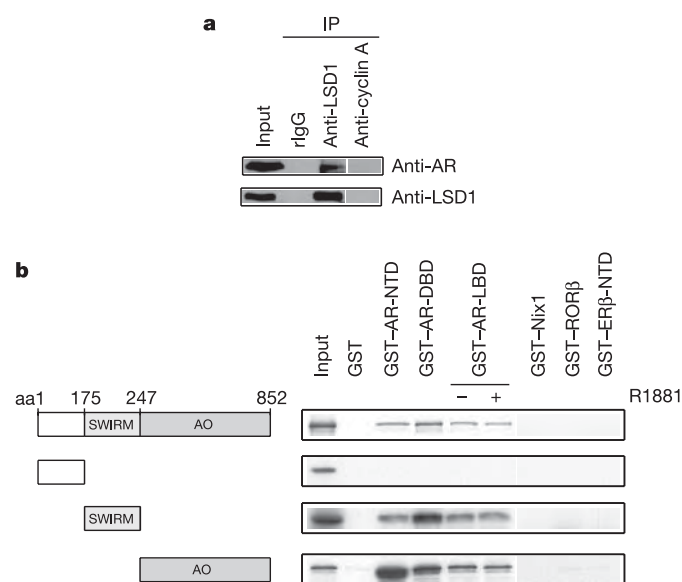


Figure 1 | LSD1 interacts with AR *in vivo* and *in vitro*. **a**, AR co-immunoprecipitates with LSD1. Extracts from mouse testis were immunoprecipitated (IP) with anti-LSD1 or anti-cyclin A antibodies and rabbit IgG as control. Western blots were decorated with anti-AR and anti-LSD1 antibodies. **b**, GST pull-down assays were performed with labelled LSD1 or LSD1 mutants and the corresponding bacterially expressed GST–AR fusion proteins. GST, GST–Nix1, GST–ROR β and GST–ER β -NTD proteins were used as controls. aa, amino acids; NTD, N-terminal domain; DBD, DNA-binding domain; LBD, ligand-binding domain.

¹Universitäts-Frauenklinik und Zentrum für Klinische Forschung, Klinikum der Universität Freiburg, Breisacherstrasse 66, 79106 Freiburg, Germany. ²Max-Planck-Institut für Immunbiologie, Stübeweg 51, 79108 Freiburg, Germany. ³Friedrich Miescher Institute for Biomedical Research, Novartis Research Foundation, Maulbeerstrasse 66, 4058 Basel, Switzerland. ⁴Institut für Pathologie, Universitätsklinikum Bonn, Sigmund-Freud-Strasse 25, 53127 Bonn, Germany.

*These authors contributed equally to this work.

demonstrate that LSD1 is a nuclear protein that co-localizes with AR in androgen-sensitive tissues such as prostate.

Because LSD1 associates with chromatin and demethylates histone H3 at lysine 4 (H3-K4) *in vitro*³, we verified that LSD1 binds to core histones, histone H3 and the N-terminal tail of histone H3 *in vitro* (Supplementary Fig. S3). To determine whether LSD1 and AR associate with chromatin *in vivo*, LNCaP cells treated with or without the synthetic AR agonist R1881 were subjected to chromatin immunoprecipitation (ChIP). Genomic DNA corresponding to the androgen response elements ARE I + II and ARE III, located in the promoter and enhancer of the prostate-specific antigen (*PSA*) gene, respectively, was immunoprecipitated in a ligand-dependent manner with anti-AR antibodies (Fig. 3a). DNA from a region between the enhancer and promoter was not enriched, thus demonstrating specificity (Fig. 3a). LSD1 specifically associates with chromatin on the *PSA* promoter both in the presence or absence of ligand (Fig. 3a). Association of LSD1 with the chromatinized *PSA* promoter is specific because DNA from neither exon 4 of the *PSA* gene nor the promoters of the *GAPDH* and *U6* genes is enriched (Fig. 3a). To demonstrate that LSD1 and AR form ligand-dependent complexes on chromatinized AREs, agonist-treated LNCaP cells were subjected to sequential chromatin immunoprecipitation (Re-ChIP), first with an anti-AR antibody and next with either anti-LSD1 antibody or rabbit IgG. Both ARE-containing regions were enriched, demonstrating that LSD1 and AR form a ligand-dependent complex on chromatin (Fig. 3a).

Because AR induces *PSA* gene expression we analysed methylation levels of repressive histone marks such as histone 3 at lysine 9

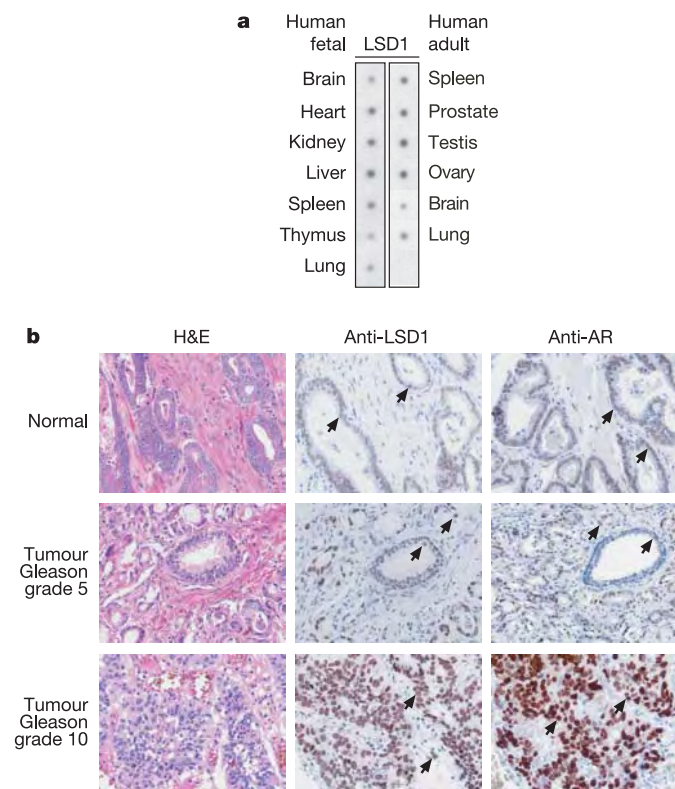


Figure 2 | LSD1 expression analyses. **a**, Expression of *LSD1* mRNA in human tissues was examined by northern blot analyses on a human multiple tissue expression array. **b**, Immunohistochemical staining of LSD1 and AR in human normal prostate and tumour prostate. LSD1 (middle column) and AR (right column) immunoreactivity is detected in the secretory epithelium of normal prostate (arrows in top row) and tumour cells (arrows in middle and bottom rows). Haematoxylin and eosin (H&E)-stained sections are shown (first column). All sections were taken from the same radical prostatectomy specimen. Magnification: $\times 250$.

(H3-K9), histone 3 at lysine 27 (H3-K27) and histone 4 at lysine 20 (H4-K20). Stimulation of LNCaP cells with R1881 results in androgen-induced transcription and is accompanied by a robust decrease in mono-, di- and trimethyl H3-K9 at the *PSA* promoter (Fig. 3b). In addition, we observed a ligand-dependent decrease in dimethyl H4-K20, whereas mono- and trimethyl H4-K20 and

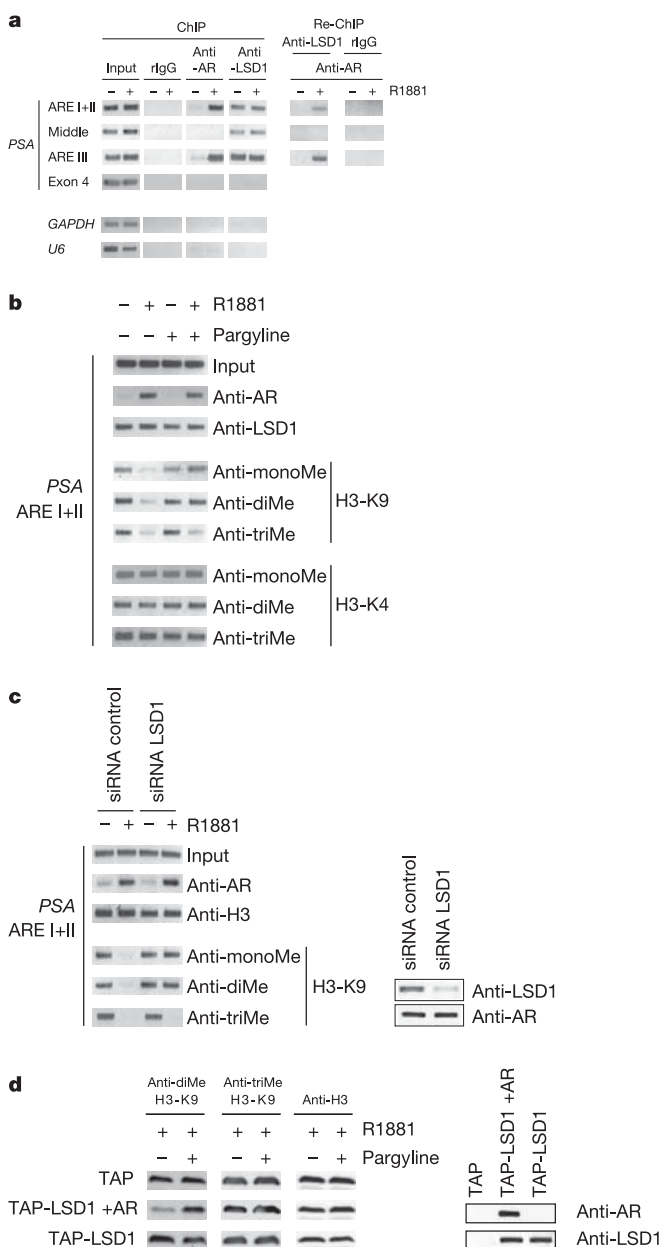


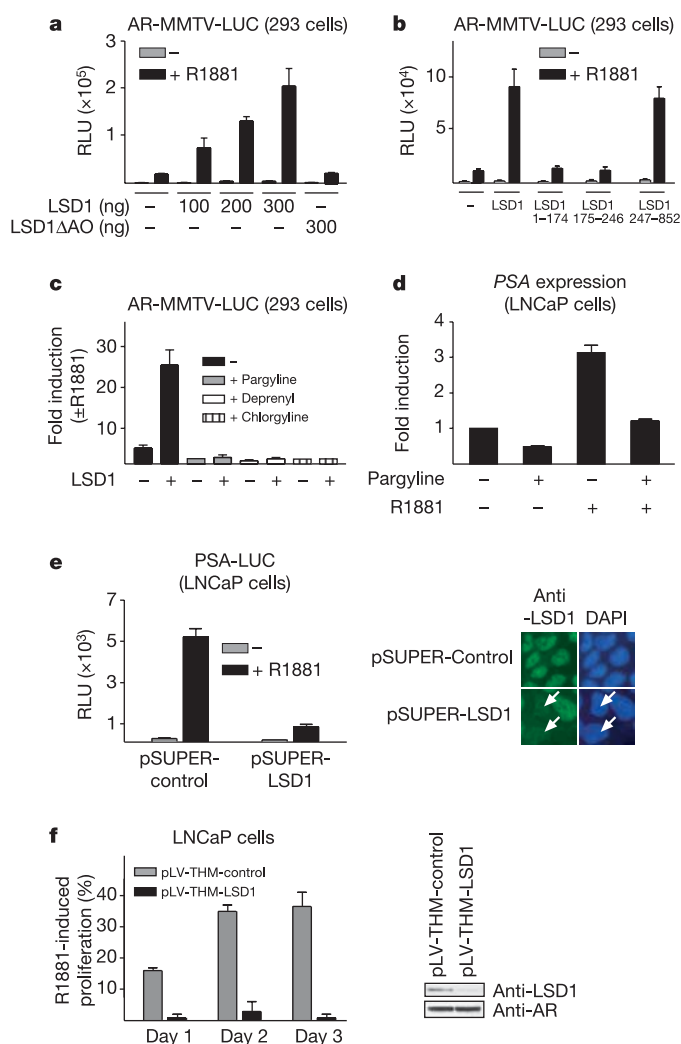
Figure 3 | LSD1 interacts with chromatin. **a–c**, LNCaP cells were incubated with or without R1881, treated with or without pargyline (**b**), or transfected with siRNA (**c**). ChIP or Re-ChIP was performed with the indicated antibodies. The precipitated chromatin was amplified by PCR using primers flanking the promoter region (ARE I + II), the middle region (middle), the enhancer region (ARE III), exon 4 of the *PSA* gene, or the promoters of the *GAPDH* and *U6* genes. siRNA-mediated knockdown of LSD1 is verified by western blot analysis (**c**, right panel) using anti-AR and anti-LSD1 antibodies. **d**, Native nucleosomes from HeLa cells were incubated in the presence of R1881 with either purified TAP, TAP-LSD1-AR, or TAP-LSD1 complexes with or without pargyline. Western blots were decorated with the indicated antibodies (left panel). The presence of LSD1 and AR in the TAP purified protein complexes was verified by western blotting using anti-AR and anti-LSD1 antibodies (right panel).

methylation levels of H3-K27 remain unchanged (Supplementary Fig. S4). Because LSD1 is an amine oxidase that catalyses demethylation, we tested whether monoamine oxidase inhibitors such as pargyline might block demethylation. Pargyline blocks demethylation of mono- and dimethyl H3-K9 during androgen-induced transcription, whereas methylation levels of trimethyl H3-K9 and the methylation status of H4-K20 and H3-K27 remain unchanged (Fig. 3b; see also Supplementary Fig. S4). Interestingly, methylation of histone H3-K4 is not altered in the presence of R1881 and not influenced by pargyline *in vivo* (Fig. 3b). To prove that LSD1 executed the ligand-dependent demethylation of mono- and dimethyl H3-K9, we designed various short interfering (si)RNAs directed against LSD1 or an unrelated control (Supplementary Fig. S5). Transfection of LNCaP cells leads to efficient and specific

downregulation of endogenous LSD1 but does not affect the level of endogenous AR (Fig. 3c). LSD1 knockdown blocks ligand-dependent demethylation of mono- and dimethyl H3-K9 but not that of trimethyl H3-K9 (Fig. 3c). The amount of total H3 on the PSA promoter is not influenced by LSD1 knockdown (Fig. 3c). To validate further that a LSD1-AR complex removes H3-K9 dimethyl marks in the presence of R1881 we established a demethylation assay *in vitro*. Tandem affinity purified (TAP) LSD1 in the presence or absence of AR (Fig. 3d) was incubated in the presence of R1881, with HeLa nucleosomes as substrate. The TAP-LSD1-AR complex demethylated dimethyl H3-K9 *in vitro*, whereas TAP-LSD1 or the TAP control failed to do so. The methylation status of the trimethyl H3-K9 control is not altered (Fig. 3d). Notably, addition of pargyline blocked demethylation of dimethyl H3-K9 by the TAP-tagged LSD1-AR complex (Fig. 3d). Thus, the *in vitro* assay demonstrates that the LSD1-AR complex directly and specifically demethylates H3-K9 and demethylation is blocked by pargyline. Taken together, these data show the ligand-dependent association of LSD1 and AR on chromatinized AREs at the promoter of the PSA gene, and the specific demethylation of the repressive histone marks mono- and dimethyl H3-K9.

Next, we performed transient transfection assays to test whether LSD1 modulates the transcriptional activity of AR. Co-expression of LSD1 and AR results in a strong ligand-dependent activation of a mouse mammary tumour virus (MMTV)-luciferase reporter (Fig. 4a), which is not observed with a LSD1 deletion mutant lacking the amine oxidase domain (LSD1 Δ AO) or in the absence of either ligand or AR (Fig. 4a; see also Supplementary Fig. S6). Stimulation of AR activity by LSD1 is potent in different cell lines, and AR-responsive minimal, synthetic and complex promoters are activated by LSD1 in a ligand-dependent manner (Supplementary Fig. S6b, c). LSD1 does not affect the transcriptional activity of related steroid hormone receptors, indicating that stimulation of AR is selective (Supplementary Fig. S6h). Furthermore, we demonstrate that the amine oxidase domain (LSD1 247–852) of LSD1 suffices to stimulate AR- and ligand-dependent reporter gene activity (Fig. 4b; see also Supplementary Fig. S6i). Because displacement of repressive histone marks by LSD1 increases AR-dependent gene expression, inhibition of LSD1 should reduce AR activity. Consequently, monoamine oxidase inhibitors such as pargyline, chlorgyline and deprenyl severely impair LSD1-induced activation of AR (Fig. 4c). In LNCaP cells, which express endogenous AR, only androgen-dependent but not unrelated reporters such as TK-LUC are inhibited by pargyline, thus demonstrating specificity (Supplementary Fig. S6j). Pargyline does not influence the activity of other nuclear receptors (Supplementary Fig. S6k). Moreover, quantitative RT-PCR analyses demonstrate that pargyline also blocks the androgen-induced expression of the endogenous PSA gene in LNCaP cells (Fig. 4d). Next, we efficiently reduced endogenously expressed LSD1 in LNCaP cells by vector (pSUPER-LSD1)-mediated RNA interference (Fig. 4e). Paralleling LSD1 knockdown, a significant ligand-dependent decrease of PSA-LUC reporter gene expression was observed (Fig. 4e), whereas expression of the unrelated TK-LUC is not influenced (data not shown). To address whether LSD1 governs androgen-dependent cell growth, we infected LNCaP cells with a lentivirus (pLV-THM-LSD1) expressing siRNA directed against LSD1. Infection with pLV-THM-LSD1 causes efficient and specific downregulation of endogenous LSD1 but does not affect the level of endogenous AR (Fig. 4f). When compared to cells transduced with the pLV-THM-control virus, androgen-induced proliferation of LNCaP cells is markedly inhibited by pLV-THM-LSD1-mediated LSD1 knockdown (Fig. 4f). These results show the physiological importance of LSD1 in the control of androgen-induced gene regulation and cell proliferation.

Our data demonstrate that AR function is controlled by the demethylase LSD1. LSD1 and AR associate at chromatinized AREs in a ligand-dependent manner, which results in concomitant specific



demethylation of the repressive histone marks mono- and dimethyl H3-K9. LSD1 has been described as a component of co-repressor complexes^{4–7}, and a recent model proposes that LSD1 represses transcription of genes silenced by Co-REST due to demethylation of the activating histone marks mono- and dimethyl H3-K4 (ref. 3). However, when in complex with AR, LSD1 demethylates the repressing histone marks mono- and dimethyl H3-K9 and thereby promotes gene activation. Thus, depending on the specific interacting partners, LSD1 action results in either gene silencing or activation. Of importance is our observation that inhibitors such as pargyline control the demethylase activity of LSD1 and thereby regulate AR. Thus, specific modulation of LSD1 activity might be a promising therapeutic target in tissues such as brain, testis and prostate, where AR has a pivotal physiological role.

METHODS

Full details of the Methods are given in the Supplementary Information.

Cell culture and transfections. 293 and CV-1 cells were cultured and transfected as described⁸. LNCaP cells were cultured in phenol-red-free RPMI1640 supplemented with 10% double-stripped fetal calf serum (dsFCS) and transfected with Effectene (Qiagen). The following amounts per well were used: 500 ng each of MMTV-LUC, ARE_{2x}-TATA-LUC, ARE_{2x}-TK-LUC, TK-LUC, TREp-LUC, β RE-LUC, ERE_{2x}-TATA-LUC, PSA-LUC, SIp-ARU-TATA-LUC; 25 ng expression plasmids for AR, PR, ER α , RAR α and TR β ; 500–700 ng expression plasmids for LSD1 1–174, LSD1 175–246, LSD1 247–852, LSD1 Δ 281–360, LSD1 Δ AO, pSUPER-control and pSUPER-LSD1; 100–700 ng expression plasmids for LSD1 were transfected per well. Chemicals were obtained as indicated: pargyline (Sigma); deprenyl and chlorgyline (ICN Biomedicals Inc.); R1881, T3, E₂, all-trans retinoic acid and R5020 (Schering AG). Cells were treated with or without 10^{-10} M R1881, 10^{-8} M R5020, 10^{-9} M E₂, 10^{-7} M T3, 10^{-6} M all-trans RA, 3×10^{-3} M pargyline, 10^{-3} M deprenyl, or 10^{-4} M chlorgyline for 18 h, as indicated. Luciferase activity was assayed as described⁹. All experiments were repeated at least five times in duplicate.

Immunohistochemistry. Polyclonal rabbit anti-LSD1 antibody was generated according to standard procedures. Stainings were performed using a protocol⁹ for antigen retrieval and indirect immunoperoxidase. Anti-AR 441 (Santa Cruz) and anti-LSD1 were used at a dilution of 1:75 and 1:500, rabbit IgG and mouse IgG (1:500; Dako) were used as secondary antibodies, and immunoreactions were visualized with the ABC complex diluted 1:50 in PBS (Vectastain, Vector).

Chromatin immunoprecipitation. ChIP experiments were performed essentially as described¹⁰. LNCaP cells were treated for 18 h with or without pargyline and for 210 min with or without 10^{-8} M R1881 as indicated. LNCaP cells were transfected 3 days before harvesting for ChIP with or without siRNA (Qiagen) following the manufacturer's instructions. Immunoprecipitation was performed with specific antibodies (anti-monoMeK9H3, anti-diMeK9H3, anti-triMeK9H3, anti-monoMeK4H3, anti-diMeK4H3, anti-triMeK4H3, anti-monoMeK27H3, anti-diMeK27H3, anti-triMeK27H3, anti-monoMeK20H4, anti-diMeK20H4, anti-triMeK20H4, anti-H3 (all obtained from Upstate Biotechnology), anti-LSD1 and anti-AR PG21 (Upstate Biotechnology)) on GammaBind-Sepharose 4B (GE-Healthcare). For PCR, 1–5 μ l out of 50 μ l DNA extract was used. For Re-ChIP assays, immunoprecipitations were sequentially washed with TSE I, TSE II, buffer III (0.25 mM LiCl, 1% NP40, 1% deoxycholate, 1 mM EDTA and 10 mM Tris-HCl pH 8.1) and TE¹⁰. Complexes were eluted by incubation with 10 mM dithiothreitol at 37 °C for 30 min, diluted 50 times with dilution buffer¹⁰ followed by a second immunoprecipitation with the indicated antibody. Primer sequences were as follows: exon 4, PSA (+3909 to +4067) 5'-GTG TGTGGACCTCCATGTTATT-3' and 5'-CCACTCACCTTTCCCTCAAG-3'; middle, PSA (–2223 to –1951) 5'-TGGGTGGGTCAGGTTTGGTT-3' and 5'-TCTTCCCTGTTTCTAGTTGAGTG-3'. PCR primers for ARE I + II (PSA (–459 to –121)), ARE III (PSA (–4288 to –3922)), GAPDH and U6 have been described previously^{3,11,12}.

Co-immunoprecipitation assays and western blot analyses. Experiments were performed essentially as described¹³. Immunoprecipitations from extracts of murine testis were performed in the presence of 1×10^{-9} M R1881 with anti-LSD1 antibody, anti-cyclin A¹³ antibody, or rabbit IgG. Western blots were decorated as indicated. Anti-AR (N20, Santa Cruz) was used. Ten per cent of testis extract was loaded as input.

Cell proliferation assay. pLV-THM-control and pLV-THM-LSD1 were used to produce recombinant lentiviruses to infect LNCaP cells as described¹⁴. The infected cells were cultured for 72 h in medium supplemented with 10% dsFCS. 0.3×10^4 cells were plated in a 96-well plate with or without 10^{-7} M R1881. The cell proliferation Elisa BrdU Colorimetric Assay (Roche) was performed accord-

ing to the manufacturer's instructions. The experiments were repeated three times in quadruplet.

Quantitative RT-PCR and statistical analysis. DNaseI-treated RNA isolated using RNawiz (Ambion) was used for reverse transcription. Quantitative PCR was performed in an ABI PRISM 7700 sequence detector. Product formation was detected by incorporation of SYBR green I using ROX as a passive reference (ABgene). The expression ratios of the analysed cDNAs were related to the normalized C_p (crossing point) of the housekeeping gene GAPDH in each sample. The following primers were used: GAPDH 5'-GAAGGTGAAGGTCG GACTC-3' and 5'-GAAGATGGTGATGGGATTTC-3'; PSA 5'-CACCTGC TCGGGTGATTCTG-3' and 5'-CCACTTCCGGTAATGCACCA-3'. Statistical analysis for quantitative PCR was performed by group-wise comparison based on PCR efficiencies and the mean crossing point deviation between sample and control group using Relative Expression Software Tool¹⁵. Experiments were repeated and analysed three times.

Demethylase assay. The demethylation assay was essentially performed as described³. TAP-tagged proteins were bound to IgG-sepharose, washed and incubated in buffer 1 (50 mM Tris pH 8.5, 50 mM KCl, 5 mM MgCl, 0.5% BSA and 5% glycerol) supplemented with 10 mM ATP, 10^{-9} M R1881 with or without 1×10^{-3} M pargyline and 1 μ g of nucleosomes purified from HeLa cells¹⁶ for 6 h at 37 °C. The reaction mixture was analysed by SDS-PAGE followed by western blotting using antibodies as indicated.

Received 27 April; accepted 8 July 2005.

Published online 3 August 2005.

- Glass, C. K. & Rosenfeld, M. G. The coregulator exchange in transcriptional function of nuclear receptors. *Genes Dev.* **14**, 121–141 (2000).
- Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
- Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).
- Shi, Y. *et al.* Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* **422**, 735–738 (2003).
- Hakimi, M. A. *et al.* A candidate X-linked mental retardation gene is a component of a new family of histone deacetylase-containing complexes. *J. Biol. Chem.* **278**, 7234–7239 (2003).
- Hakimi, M. A. *et al.* A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc. Natl Acad. Sci. USA* **99**, 7420–7425 (2002).
- Eimer, S. *et al.* Loss of spr-5 bypasses the requirement for the *C. elegans* presenilin sel-12 by derepressing hop-1. *EMBO J.* **21**, 5787–5796 (2002).
- Müller, J. M. *et al.* FHL2, a novel tissue-specific coactivator of the androgen receptor. *EMBO J.* **19**, 359–369 (2000).
- Müller, J. M. *et al.* The transcriptional coactivator FHL2 transmits Rho signals from the cell membrane into the nucleus. *EMBO J.* **21**, 736–748 (2002).
- Shang, Y., Myers, M. & Brown, M. Formation of the androgen receptor transcription complex. *Mol. Cell* **9**, 601–610 (2002).
- Shatkina, L. *et al.* The cochaperone Bag-1L enhances androgen receptor action via interaction with the NH₂-terminal region of the receptor. *Mol. Cell. Biol.* **23**, 7189–7197 (2003).
- Kang, Z., Pirskanen, A., Jänne, O. A. & Palmivo, J. J. Involvement of proteasome in the dynamic assembly of the androgen receptor transcription complex. *J. Biol. Chem.* **277**, 48366–48371 (2002).
- Metzger, E. *et al.* A novel inducible transactivation domain in the androgen receptor: implications for PRK in prostate cancer. *EMBO J.* **22**, 270–280 (2003).
- Wiznerowicz, M. & Trono, D. Conditional suppression of cellular genes: lentivirus vector-mediated drug-inducible RNA interference. *J. Virol.* **77**, 8957–8961 (2003).
- Pfaffl, M. W., Horgan, G. W. & Dempfle, L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* **30**, e36 (2002).
- O'Neill, T. E., Roberge, M. & Bradbury, E. M. Nucleosome arrays inhibit both initiation and elongation of transcripts by bacteriophage T7 RNA polymerase. *J. Mol. Biol.* **223**, 67–78 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Benzing, F. Claessens, T. Jenuwein, Z. Sun and D. Trono for providing reagents. We are obliged to the members of the Schüle laboratory for discussions. We thank K. Fischer, P. Kahl and L. Heukamp for technical assistance. This work was supported by grants from the Deutsche Forschungsgemeinschaft and Deutsche Krebshilfe to R.S.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.S. (roland.schuele@uniklinik-freiburg.de).

LETTERS

Mesoscale conformational changes in the DNA-repair complex Rad50/Mre11/Nbs1 upon binding DNA

Fernando Moreno-Herrero¹, Martijn de Jager², Nynke H. Dekker¹, Roland Kanaar^{2,3}, Claire Wyman^{2,3} & Cees Dekker¹

The human Rad50/Mre11/Nbs1 complex (hR/M/N) functions as an essential guardian of genome integrity by directing the proper processing of DNA ends, including DNA breaks¹. This biological function results from its ability to tether broken DNA molecules^{2,3}. hR/M/N's dynamic molecular architecture consists of a globular DNA-binding domain from which two 50-nm-long coiled coils protrude. The coiled coils are flexible⁴ and their apices can self-associate⁵. The flexibility of the coiled coils allows their apices to adopt an orientation favourable for interaction. However, this also allows interaction between the tips of two coiled coils within the same complex, which competes with and frustrates the inter-complex interaction required for DNA tethering. Here we show that the dynamic architecture of hR/M/N is markedly affected by DNA binding. DNA binding by the hR/M/N globular domain leads to parallel orientation of the coiled coils; this prevents intracomplex interactions and favours intercomplex associations needed for DNA tethering. The hR/M/N complex thus is an example of a biological nanomachine in which binding to its ligand, in this case DNA, affects the functional conformation of a domain located 50 nm distant.

The hR/M/N complex has essential functions in various aspects of genome metabolism that involve the processing of DNA ends, such as homologous recombination, non-homologous end joining and maintenance of telomere length¹. The molecular mechanism underlying these biological functions of hR/M/N involves tethering DNA molecules by means of the interaction between DNA-bound hR/M oligomers^{2,3}. The hR/M complex is a heterotetramer, R₂M₂, arranged with a globular DNA-binding domain, including the Mre11 dimer and the two Rad50 ATPase domains, from which the long intramolecular coiled coils of Rad50 protrude^{2,6,7}. The coiled-coil apex contains a CXXC amino-acid motif that forms a structure described as a zinc hook. Genetic experiments in *Saccharomyces cerevisiae* have shown that the zinc hook is an important determinant of Rad50 function³. Two CXXC motifs can dimerize by the coordination of a Zn²⁺ ion, providing a possible interface for interaction between hR/M complexes⁵. hR/M complexes form oligomers on linear DNA where interactions between the apices of the coiled coils then tether DNA molecules⁸. Conformational changes that alter the orientation, flexibility or dynamics of the coiled coils could be exploited to control intercomplex versus intracomplex interaction of the coiled-coil apices and therefore the biological function of hR/M. Here we resolve this issue by determining the influence of nucleotide cofactor and DNA binding on the architectural dynamics of hR/M and hR/M/N by using time-resolved atomic force microscopy (AFM) in solution.

DNA tethering requires only the hR/M complex *in vitro*. We therefore first analysed the dynamic architecture of this form, which is relevant for activity. The hR/M complex globular domain was 6.4 ± 1.0 nm high, and the coiled coils were 50 ± 3 nm long and 2.0 ± 0.3 nm high in buffer ($n = 30$) (see Fig. 1). In these high-resolution images of hR/M complexes, two, three and sometimes four individual domains could be resolved inside the globular part of the protein (Fig. 1C–f). These are likely to be the separate Mre11 molecules and Rad50 ATPase domains. The coiled coils of individual hR/M complexes adopted either an open conformation in which the apices of the coiled coils did not touch each other (Fig. 1A) or a closed conformation in which the apices of the coiled coils were in contact (Fig. 1B)^{2,9}. In the closed conformation, the junction of the coiled coils was notably higher, 3.9 ± 0.5 nm, than that of an individual coiled coil. Single complexes frequently switched between open and

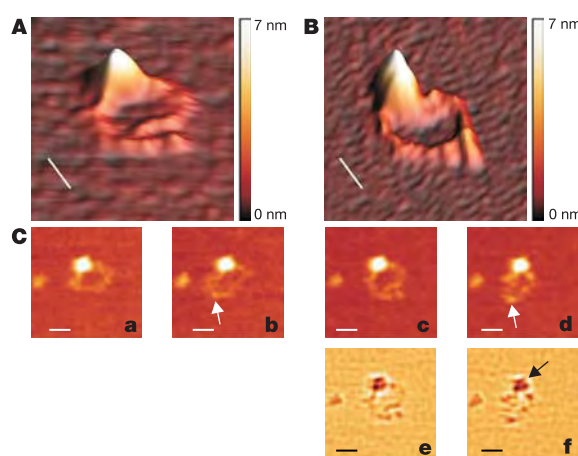


Figure 1 | Time-resolved high-resolution AFM images of the hR/M complex in buffer. **A, B,** From the globular domain, two coiled coils extend that are either not connected (open conformation, **A**) or overlapping near the apices (closed conformation, **B**). **C,** Opening and closing of the coiled coils is a dynamic process, as can be seen in the four-image sequence extracted from a movie (39 s per frame) of 13 min (**a–d**). The closed coiled-coils conformation is very often associated with the presence of a distinct point at the coiled-coil junction (white arrows). The frames in **e** and **f** are height images filtered with a laplacian operator to enhance substructure details. Two, three and occasionally four separate areas can be resolved inside the globular domain of hR/M (black arrow). Scale bars, 25 nm.

¹Kavli Institute of Nanoscience, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands. ²Department of Cell Biology and Genetics and ³Department of Radiation Oncology, Erasmus MC, PO Box 1738, 3000 DR Rotterdam, The Netherlands.

closed conformations as a function of time. In Fig. 1Ca–d, four consecutive images (39 s per image) show two switches between open and closed conformations (see also Supplementary Movie 1). The interaction between the apices of the coiled coils was not primarily due to disulphide bridges between exposed cysteine residues, because the addition of 2 mM dithiothreitol (DTT) to the imaging buffer did not affect the proportion of complexes with the closed conformation (48% ($n = 232$) and 45% ($n = 50$) with and without DTT, respectively).

The orientation of the hR/M coiled coils was determined in the presence and absence of the non-hydrolysable ATP analogue β - γ -imidoadenosine 5'-phosphate (AMP-PNP) and Mg^{2+} . This was quantified by measuring the angle α between the coiled coils and the longitudinal axis of the complex (see Fig. 2a). In addition, a value d for the distance between the coiled coils at the junction with the globular domain was measured (Fig. 2a). The standard deviation of the distribution of α for each molecule, displayed as error bars in Fig. 2b, provides information on the degree of mobility of the coiled coils. Molecules were analysed in three conditions (Fig. 2b): in the absence of nucleotide cofactor (black circles); after incubation with AMP-PNP but imaged in a buffer without AMP-PNP (green squares); and after incubation with AMP-PNP and imaged in the presence of AMP-PNP (blue triangles). The mean values for α obtained under these three conditions were $27 \pm 8^\circ$, $26 \pm 9^\circ$ and $24 \pm 8^\circ$, respectively, indicating that nucleotide cofactor binding does not influence the orientation or the flexibility of the coiled coils. This was unexpected, given that binding AMP-PNP causes a 30° rotation of the amino-terminal domain with respect to the carboxy-terminal domain of the *Pyrococcus furiosus* Rad50 ATPase and that the coiled coils connect these two domains⁶. If the same conformational change occurs within the ATPase domain of hRad50 after nucleotide cofactor binding, it is apparently not transmitted to the attached coiled coils. There was no correlation of the value of α with the conformation of the coiled coils: for the open conformation α was $24 \pm 8^\circ$ (52%, or 120 of 232), for the closed conformation α was $27 \pm 8^\circ$ (48%, or 112 of 232). Correspondingly, the mean value for d , 14 ± 3 nm, did not vary significantly between the molecules in the three conditions used (Fig. 2c).

Binding of hR/M to DNA induced a strikingly different arrangement of the coiled coils compared with unbound hR/M complexes. When bound to DNA by means of the globular domain, both coiled coils of the complex were parallel and on the same side of the DNA. Figure 3A shows a gallery of images of hR/M complexes bound to circular DNA in which most hR/M (93%, $n = 15$) had the architecture described above. The conformational change in hR/M was also clear in the presence of 90-base-pair (bp) linear DNA (87% parallel

conformation, $n = 44$) (Fig. 3B). Movies of DNA-bound hR/M complexes showed that the coiled coils retain some flexibility and that they move synchronously. The characteristic high point located at the junction of the coiled coils (Fig. 1B) was not observed for any of the hR/M complexes bound to DNA, indicating that within a DNA-bound complex the apices of the coiled coils do not overlap. For the DNA-bound complexes, the mean angle α was $6 \pm 5^\circ$ (Fig. 3Da) and the mean distance d was 8 ± 3 nm. Thus, α is much smaller than the $26 \pm 9^\circ$ found for DNA-free hR/M. The strong effect of DNA binding on coiled-coil orientation is summarized in the histogram of the distributions of α for DNA-bound and DNA-unbound hR/M (Fig. 3Db). The presence of AMP-PNP and Mg^{2+} did not change the architecture of DNA-bound hR/M (Fig. 3C).

Time-resolved imaging in buffer also revealed the dynamic transition between the architecture of DNA-bound hR/M and the architecture of DNA-free hR/M. Figure 3Fa–f shows part of a longer movie (Supplementary Movie 2) in which, first, an hR/M complex was bound to DNA (for 8 min) and then, between the third and fourth frames, it dissociated from the DNA. Simultaneously, it changed from the parallel to the bent, open and apices-joined conformation (Fig. 3Fc, d shows the key transition frames, enlarged in Fig. 3E, G). On several occasions, the hR/M complexes were observed to move along the DNA while following the contour of the DNA molecule (Supplementary Fig. 1). This shows that hR/M can slide along DNA rather than transfer between different DNA segments. Oligomers of hR/M form only on linear DNA but are not always found at DNA ends¹⁰. This sliding mechanism could therefore serve to clear DNA ends for further repair processing.

The hR/M complex can also include a third subunit, Nbs1. The Nbs1 component probably interacts with the globular domain of the R/M complex and functions in signalling the presence of DNA breaks to the cell cycle checkpoint machinery¹¹. The three-component complex hR/M/N appeared very similar to hR/M in our AFM study (Fig. 4a). The presence of the third subunit was evident as a larger globular domain with an average height of 8.0 ± 0.8 nm (Fig. 4b). In the absence of DNA, the hR/M/N complex also had bent, open coiled coils with apices in open and closed conformations. After DNA binding, in the presence of a molar excess of a 90-bp oligonucleotide, the coiled coils of hR/M/N also adopted a parallel conformation (86%, $n = 85$; Fig. 4c, d). The use of short DNA oligonucleotides for binding prevented the occurrence of large hR/M/N oligomers and thus allowed us to observe single intercomplex interactions through the apices of the coiled coils. Indeed, as shown in Fig. 4e, the intercomplex interaction of coiled-coil apices necessary for tethering was occasionally observed in these conditions (about 10% of the complexes observed). We never observed such an

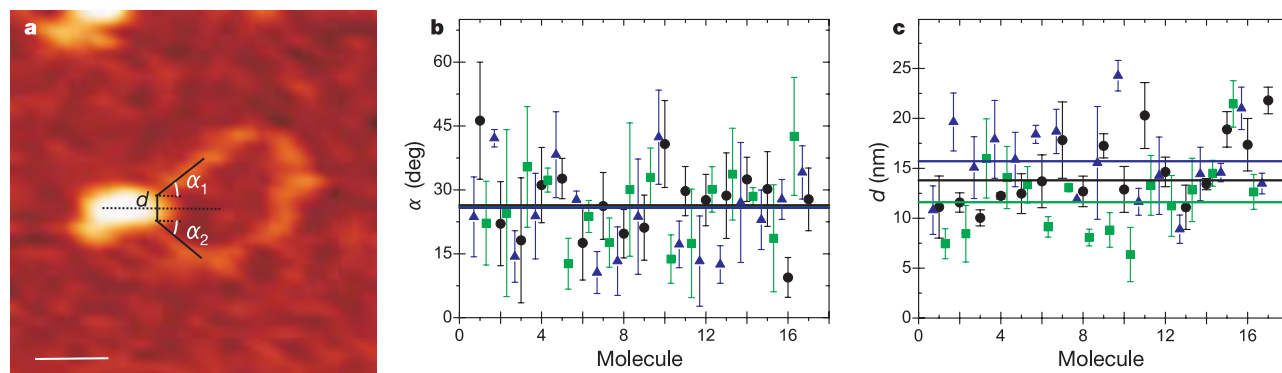


Figure 2 | Effect of nucleotide cofactor binding on the orientation of hR/M coiled coils. **a**, Definition of angle α ($= (\alpha_1 + \alpha_2)/2$) and distance d . Movies composed of 6–20 frames were acquired for 17 different hR/M molecules. **b**, **c**, Values (means $\pm$ s.d.) of α and d obtained for each molecule. Black circles show the results for hR/M in the absence of AMP-

PNP; green squares show the results when the hR/M complex was incubated for 15 min in the presence of 2 mM $MgCl_2$ and 2 mM AMP-PNP; blue triangles show the results when, after a similar incubation, 2 mM AMP-PNP was also present in the imaging buffer.

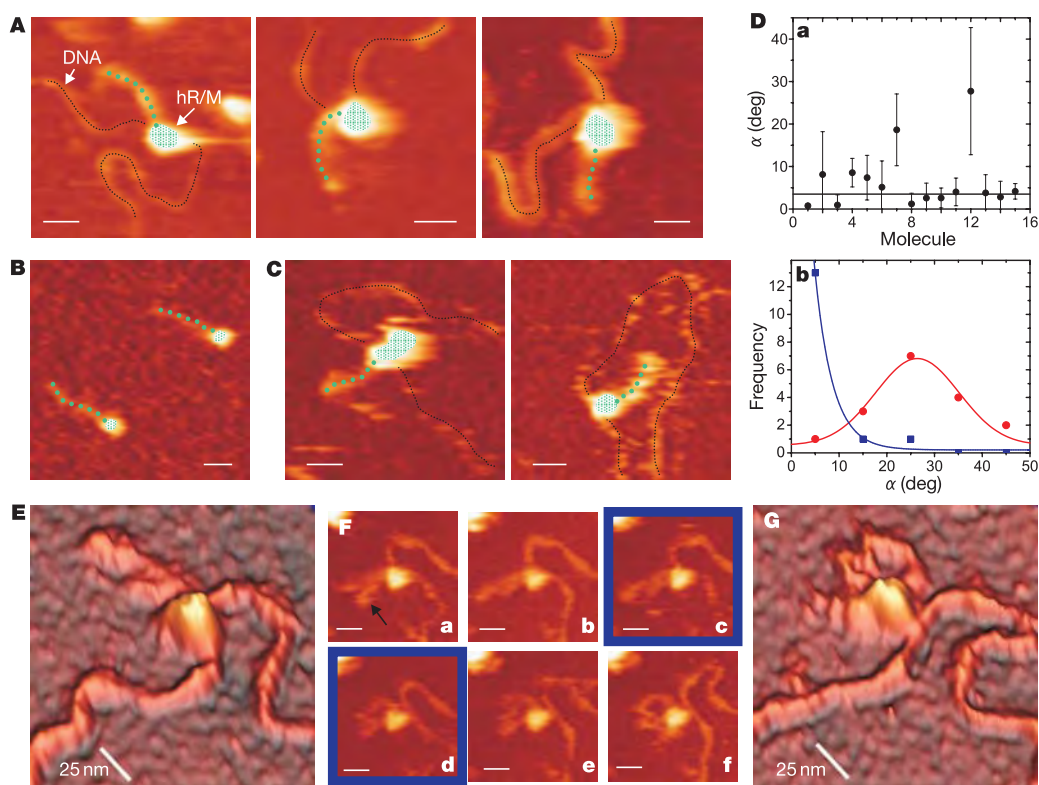


Figure 3 | hR/M binding to DNA induces a parallel alignment of the coiled coils. **A, B,** AFM images of hR/M complexes bound to circular double-stranded DNA (dsDNA) (**A**) and incubated with a threefold to sixfold molar excess of a 90-bp linear dsDNA (**B**). The positions of DNA and hR/M are indicated, to guide the eye. A single, thick, coiled coil protrudes from the hR/M globular domain that is bound to the DNA. **C,** A similar result was obtained when AMP-PNP was present in the binding reaction. **D,** Values (means $\pm$ s.d.) of angle α for DNA-bound hR/M (**a**) and histogram of the distributions of α for DNA-bound (blue curve) and DNA-free (red curve) hR/M (**b**). Solid lines are fits to the data. **F, a–f,** Six consecutive images extracted from a movie (39 s per frame) showing the dynamic transition from DNA-bound to free hR/M. Occasionally, the apices of the coiled coils separate from each other (black arrow). The transition frames (**c** and **d**) are highlighted and displayed in a three-dimensional rendering in **E** and **G**, enlarged and rotated through 60° . Scale bars, 30 nm unless otherwise labelled.

intercomplex configuration for hR/M and hR/M/N in the absence of DNA.

The existing molecular picture of hR/M/N function involves three processes: first, binding of the complex to DNA ends; second, the formation of large DNA-bound oligomers with the Rad50 apices of the coiled coils protruding; and third, subsequent interaction between the coiled coils of hR/M/N complexes bound to different DNA ends, to tether them and keep them close in nuclear space^{8,11}.

Interaction between the apices of the coiled coils is therefore essential for R/M/N function³. A conundrum of this molecular description centres on the control of this interaction. A single hR/M/N complex has two such apices in high local concentration. Because of the flexibility of the Rad50 coiled coils these can, and do, interact. DNA-bound hR/M/N oligomers also have many apices in high local concentration, and interaction between those bound to the same DNA molecule would be futile for tethering independent DNA ends.

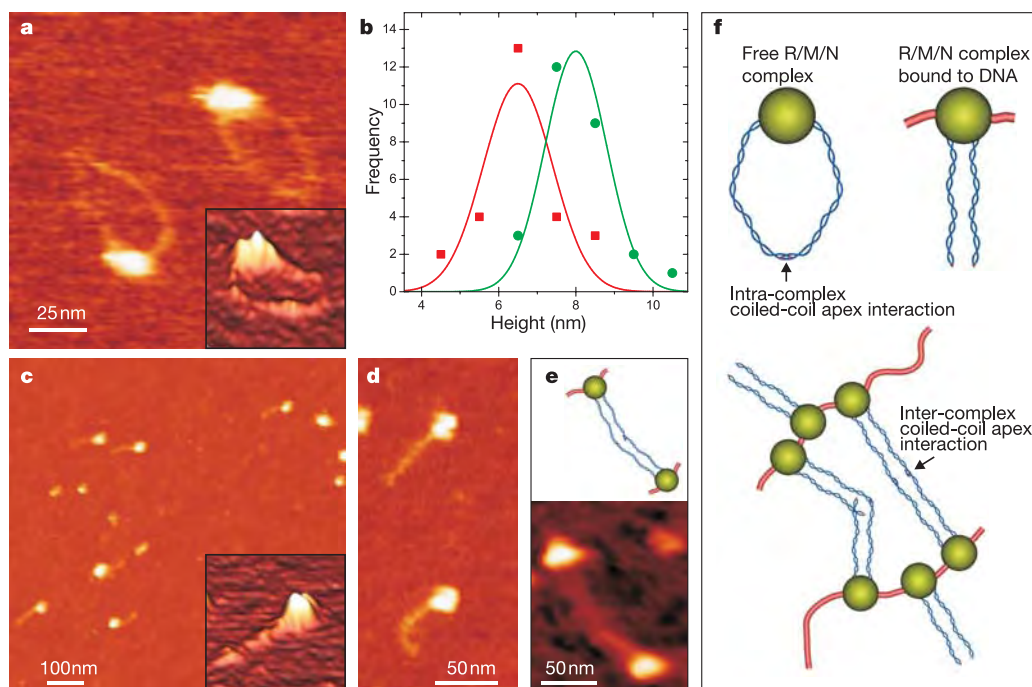


Figure 4 | Mesoscale conformational change in hR/M/N upon DNA binding and its implications for DNA-end tethering. **a,** AFM images of free hR/M/N show an architecture similar to that of hR/M. **b,** Histogram of heights for hR/M/N (green) and hR/M (red) globular domains. Solid lines are fits to the data. **c,** In the presence of a molar excess of 90-bp dsDNA, 87% of hR/M/N complexes showed a single, thick coiled coil. Insets in **a** and **c** show three-dimensional renderings of free and DNA-bound hR/M/N complexes. **d,** Detailed image showing the striking architecture of DNA-bound hR/M/N. **e,** Example of intercomplex coiled-coil apex interactions. **f,** Diagram showing the importance of conformational changes in the hR/M or hR/M/N coiled coils for DNA-end tethering.

Here we have shown that this futile intracomplex joining of Rad50 coiled-coil apices is prevented once DNA is engaged. The hR/M/N coiled coils become oriented in such a way that intracomplex interactions are prevented, while at the same time intercomplex interactions needed for tethering are favoured (Fig. 4f).

METHODS

Protein preparation. hR/M and hR/M/N were purified as described². Stock solutions ($1 \mu\text{g} \mu\text{l}^{-1}$ in 150 mM KCl, 2 mM DTT, 25 mM Tris-HCl pH 8.0, 10% glycerol) were divided into aliquots and stored at -80°C . A fresh aliquot was used for each AFM experiment.

DNA. pGEM-3Z plasmid (2,743 bp) was relaxed by using the nicking enzyme N.BstNB I (New England Biolabs). The isolated product was checked by gel electrophoresis and AFM inspection. A DNA fragment of 90 bp and four nucleotide overhangs was generated by annealing two complementary 94-mers, checked by gel electrophoresis and isolated with a DNA purification kit (MoBio Laboratories).

AFM sample preparation. For each experiment a protein aliquot was thawed, diluted 1:50 in storage buffer and kept on ice. A $20 \mu\text{l}$ portion of the diluted solution (400 ng of protein complex in 150 mM KCl, 25 mM Tris-HCl pH 8.0, 10% glycerol) was placed on a freshly cleaved mica disc held inside the liquid cell. After 2 min the liquid cell was filled with 1 ml of imaging buffer (3 mM MgCl_2 , 75 mM KCl, 25 mM Tris-HCl pH 8.0). The sample was never allowed to dry. Imaging buffer conditions were carefully optimized to allow movement of the molecules on the surface while maintaining the best possible image quality. The use of large volumes of imaging buffer minimized any variation in potassium and magnesium concentrations due to evaporation. To study the effect of nucleotide cofactor on the conformation of the coiled coils, the protein complex was incubated for 15 min at room temperature ($\sim 22^\circ\text{C}$) in 150 mM KCl, 25 mM Tris-HCl pH 8.0, 10% glycerol, containing 2 mM AMP-PNP and 2 mM MgCl_2 . Subsequent sample preparation proceeded as described above.

DNA binding reaction. hR/M DNA binding reactions included 10 ng of circular DNA and 100 ng of hR/M in 15 mM MgCl_2 , 75 mM KCl, 15 mM Tris-HCl pH 8.0, 5% glycerol in a total volume of $10 \mu\text{l}$. The binding reaction was performed for 15 min at room temperature. The whole reaction mixture was deposited on a mica disc and imaging was done in 1 ml of imaging buffer supplemented with 2 mM hR/M. hR/M and hR/M/N were mixed with the 90-bp oligonucleotide in the same conditions with a twofold to eightfold molar excess of DNA over protein.

AFM imaging. hR/M and hR/M/N DNA binding reactions were performed in solution before the sample was deposited on a mica surface. Biological particles were partly immobilized on mica by carefully tuning the AFM imaging conditions to allow the detection of movement. Samples were imaged in buffer with a commercial AFM from Nanotec Electronica operating in tapping mode, using Olympus cantilevers (0.05 N m^{-1} , Olympus TR400 PSA). Typical tapping amplitudes at imaging were 5–7 nm at the resonance frequency of the cantilevers in liquids (9.5–10.5 kHz). Images were acquired at rates of 30 s per image (128 pixels, 500 nm) and 60 s per image (256 pixels, 1,000 nm). Imaging was always from left to right (fast scan direction) and from up to down (slow scanning direction), to minimize any piezo creep effects. The absence of any correlation between the scanning direction and the orientation of the hR/M coiled coils as well as the observation of objects moving in the direction opposite to scanning indicate that the observed dynamics were not dominated by interaction of the scanning tip with the sample.

Image processing and measurements. Image processing was performed with WSxM software (www.nanotec.es). Standard image processing consisted of

plane subtraction and flattening. To capture dynamic events, consecutive images were taken at the same spot. These images were aligned by using cross-correlation filters and shown sequentially to build up a movie. By checking consecutive frames it was possible to observe which parts of the hR/M complex or which areas of DNA were mobile and loosely bound to the mica surface. The measurement of the angle between hR/M coiled coils was done by manually tracing the direction of both coiled coils in the vicinity of the globular domain. The contact of the traced line with the globular domain defines a segmented line with two angles, α_1 and α_2 , and a distance d . α is defined as the arithmetic mean of α_1 and α_2 .

Received 21 January; accepted 19 June 2005.

1. Connelly, J. C. & Leach, D. R. Tethering on the brink: the evolutionarily conserved Mre11–Rad50 complex. *Trends Biochem. Sci.* **27**, 410–418 (2002).
2. de Jager, M. *et al.* Human Rad50/Mre11 is a flexible complex that can tether DNA ends. *Mol. Cell* **8**, 1129–1135 (2001).
3. Wiltzius, J. J. W., Hohl, M., Fleming, J. C. & Petrini, J. H. J. The Rad50 hook domain is a critical determinant of Mre11 complex functions. *Nature Struct. Mol. Biol.* **12**, 403–407 (2005).
4. van Noort, J. *et al.* The coiled-coil of the human Rad50 DNA repair protein contains specific segments of increased flexibility. *Proc. Natl Acad. Sci. USA* **100**, 7581–7586 (2003).
5. Hopfner, K. P. *et al.* The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. *Nature* **418**, 562–566 (2002).
6. Hopfner, K. P. *et al.* Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* **101**, 789–800 (2000).
7. Hopfner, K. P. *et al.* Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50 ATPase. *Cell* **105**, 473–485 (2001).
8. Wyman, C. & Kanaar, R. Chromosome organization: reaching out to embrace new models. *Curr. Biol.* **12**, R446–R448 (2002).
9. de Jager, M. *et al.* Differential arrangements of conserved building blocks among homologs of the Rad50/Mre11 DNA repair protein complex. *J. Mol. Biol.* **339**, 937–949 (2004).
10. de Jager, M. *et al.* DNA-binding and strand-annealing activities of human Mre11: implications for its roles in DNA double-strand break repair pathways. *Nucleic Acids Res.* **29**, 1317–1325 (2001).
11. Stracker, T. H., Theunissen, J. W., Morales, M. & Petrini, J. H. J. The Mre11 complex and the metabolism of chromosome breaks: the importance of communicating and holding things together. *DNA Repair (Amst.)* **3**, 845–854 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Paull for the gift of the baculoviruses producing hRad50, hMre11 and hNbs1, and R. Seidel for useful discussions. F.M.-H. is supported by a postdoctoral fellowship from La Fundación Ramón Areces. M.d.J. is supported by an EUR fellowship from the Erasmus MC. This project is supported in part by a grant from NWO-FOM/ALW (Netherlands Organization for Scientific Research) to R.K., C.W. and C.D. Work in the laboratories of R.K. and C.W. is supported by grants from the European Commission, NWO and the Dutch Cancer Society. Work in the laboratory of C.D. and N.D. acknowledges support from FOM and NWO.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.D. (dekker@mb.tn.tudelft.nl) and C.W. (c.wyman@erasmusmc.nl).

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

Time for a change

Tradition! That's the chorus and title for the opening song of *Fiddler on the Roof*, a musical set in a Jewish *shtetl* in 1905 Russia. The play refers to the difficulties that dairyman Tevye has with accepting change. But if a recent letter to the *Journal of the American Medical Association* (JAMA) is to be believed, the play's themes also resonate with graduate medical education (K. M. Ludmerer and M. M. E. Johns *J. Am. Med. Assoc.* **294**, 1083–1087; 2005).

The letter says that graduate medical education has failed to live up to its potential partly because of a "traditional subordination of education to service". But this sentiment broadly applies to all scientific education. In medicine, it means medical residents end up working 24-hour-plus shifts with little educational support. For scientific graduate students and postdocs, it means periods of training when their progress is beholden to principal investigators, who sometimes prolong the process to get more publications out of the young scientists they are supposedly mentoring.

Both the JAMA letter and this week's Recruiters & Academia (see page 448), offer some alternatives to these

traditions. The JAMA item suggests limiting the number of hours residents spend with patients, relieving residents of non-educational tasks, improving the educational content of training, and providing a support system to ease emotional stress. The *Naturejobs* item proposes taking some emphasis off publication record and rewarding other aspects of scientific work.

These changes could make science and medicine more attractive as career options. And they could make the professional development for both MDs and PhDs more productive — especially if it means they are treated as trainees, rather than workhorses. Ideally, better science and medicine would result. In *Fiddler*, Tevye eventually accepts change. Perhaps the scientific and medical establishment should as well.



Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

US Head Office, New York
 345 Park Avenue South, 10th Floor,
 New York, NY 10010-1707
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager/Corporations:
 Peter Bless
 Classified Sales Representatives
 Tel: +1 800 989 7718

**New York/Pennsylvania/
 Latin America:** Kelly Roman
**Midwest USA/Maryland/
 NIH:** Wade Tucker
East USA/Canada:
 Janine Taormina

San Francisco Office
Classified Sales Representative:
 Michaela Bjorkman
 225 Bush Street, Suite 1453
 San Francisco, CA 94104
 Tel: +1 415 781 3803
 Fax: +1 415 781 3805
 e-mail: m.bjorkman@naturesf.com

European Head Office, London
 The Macmillan Building,
 4 Crinan Street,
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director: Nevin Bayoumi (4978)
European Sales Manager: Andy Douglas (4975)

Advertising Production Manager: Billie Franklin
 To send materials use London address above.
 Tel: +44 (0) 20 7843 4814
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
 Patrick Phelan
 e-mail: p.phelan@nature.com

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

Small steps towards campus child care

Dedicated fellowships, mentoring networks and tenure-adjustment programmes have been designed to promote women in science. But Anne Bertolotti, a geneticist at the Ecole Normale Supérieure in Paris, sees only one realistic way to meet that goal: provide day care. Tiny tots can prove formidable foes to a woman's academic career aspirations. Without affordable child-care options, long hours and low pay force many postdocs to make an unfortunate choice between work or motherhood.

Although Paris offers government-sponsored day care, demand exceeds supply. With care for one child already costing 25% of her income, Bertolotti faces typical concerns as she seeks a spot for her second child — availability, accessibility and cost.

The shortage of high-quality, affordable day care is enough to make academic parents cry. The US National Postdoctoral Association (NPA) says that, along with retirement plans and insurance, child care is one of the top three issues raised in surveys. Although on-site child-care facilities are rare at research institutions, a number of universities have recognized the recruitment, retention and employee-satisfaction benefits they offer. Unfortunately, old-fashioned attitudes and cost remain barriers at most institutions.

Home from home

The luxury of on-site day care is offered by a handful of research institutes and universities in countries such as the United States, Britain, Germany and Sweden. A few

For women researchers, child care can be a major obstacle to getting back to the lab. Virginia Gewin looks at the options for working mums.

countries, including Sweden and France, have government-run facilities. But most academics around the world rely on private facilities. More than one-third of US postdocs have children, yet only 10% of the 70 US research institutions informally surveyed by the NPA offer child-care facilities.

The waiting lists for campus facilities is a testament to their success. HutchKids, a non-profit facility at the Fred Hutchinson Cancer Center in Seattle, Washington, has 120 slots for infants and children. More than 130 kids are currently on the waiting list. "The kind of work that scientists do, especially here at the Hutch, demands that they have quality child care," says HutchKids director Nancy Myles.

Like many US universities, Princeton is taking baby steps to deal with the issue. Although Princeton has offered on-site child care for decades, a recent baby boom has increased demand. A health and well-being task force, created three years ago, estimated a need to double capacity. In the planning phase of the expansion, huge questions remain about where and how to run such facilities on campus.

The situation is similar in Britain. Although some UK universities, such as the University of Nottingham, offer non-profit day-care services, the demand for centres nearby is so high that private entities fill the gap. In the past few years, four new day-care centres have opened around the University of Cambridge.

In Japan, day care is almost exclusively private. With dozens of research institutes and universities in a city of 10 million people, some Tokyo academics have to travel for up to an hour outside the city to find a place.

With one of the most challenging child-care situations in Europe, fewer than 5% of German research institutes and universities offer day care. Those that do, such as Heidelberg's

European Molecular Biology Laboratory (EMBL) and the University of Heidelberg,

P. ROTHE/UNIV. HEIDELBERG



U. SIRBORN/KAROLINSKA INST.



Some universities and institutes see day-care units as useful benefits for recruiting good employees.

P. ROTHE/UNIV. HEIDELBERG

find that it often makes the difference in researcher recruitment. Jan Ellenberg, an EMBL cell biologist, says that the lack of a public day-care system in Germany would have made it impossible for him and his wife — both faculty members — to work. “The best scientists will have choices on the job market, and having excellent day-care facilities makes EMBL a more competitive employer,” he says.

Tough decisions

A recent survey by Germany's Center of Excellence for Women in Science suggested that more than 40% of academic women choose not to have children; many others leave academia. Although remnants of the former socialist system still provide options for women in eastern Germany, academics in the west depend on parents willing to donate the time and energy to build and run their own day-care centres. Christa Schleper, now an environmental biologist at the University of Bergen in Norway, and her husband spent countless unpaid hours maintaining such a centre while working at the Darmstadt University of Technology in Germany.

In Norway, surprisingly, Schleper faces a similar problem. Although Scandinavian governments are family-friendly, places in public day-care centres are increasingly hard to find in some Norwegian cities. In fact, the availability of child care has become a hot political issue in Norway's upcoming elections.

Government-run programmes, such as those in Scandinavia, are accessible to everyone. That is not always the case in on-campus facilities. Although many US research organizations offer on-campus child-care services, they sometimes limit access to postdocs and graduate students, giving priority to faculty members. Indeed, organizations that actively



recruit husband-and-wife research teams, such as the University of Texas Southwestern Medical Center in Dallas, use child-care centres as an enticement. “We're starting to recognize it's not just a women's issue, it's a dual-career family issue,” says Kim Orth, a biomedical researcher there.

Some facilities, such as the parent-run HutchKids, are first come, first served. Even so, high costs can make campus facilities inaccessible to postdocs and students.

Paying the price

With campus day-care centres costing roughly \$1.5 million a year to run, it is not surprising that most institutions have yet to address the issue. Indeed, cost remains the biggest hurdle — particularly in countries such as the United States and Britain. Postdocs there can spend half or more of their salaries on child care, whereas continental Europeans and Japanese are more likely to pay 10–25%, thanks to generous government or institutional subsidies.

The Hutch offers a subsidy of \$250 per month at its own (or any other nearby accredited) facility, lowering the cost from almost 50% to roughly 35% of annual salary. The NPA found only three other US institutions — the California Institute of Technology, Fox Chase Cancer Research Center and Johns Hopkins University — that offer similar subsidies. Princeton University offers need-based financial aid, rather than a straight subsidy. Qualifying families pay just 10% if combined annual incomes are less than \$50,000.

Britain has started to offer a government voucher equivalent to £100 (US\$184) per month to help lower costs. But postdocs at UK campuses can expect to pay about 30% of their salary, or £600 a month, for infant day care. The University of Nottingham has offered university-sponsored child-care services for ten years, but demand has still led to the creation of private centres on campus.

Government subsidies, particularly generous ones such as in Sweden, make a huge dent in child-care costs. Stockholm's Karolinska University Hospital offers day care for employees that is the equivalent of about 5–10% of a postdoc's salary: only 10% of the real cost. The remainder is paid by tax initiatives. Norwegians pay more for child care and the subsidies are designed to encourage women to stay at home for a longer period of time — another contentious political issue.

With government help, parent-organized day care in Germany usually ends up being inexpensive. The EMBL facility is unusual in that it is open to all faculty members and postdocs and asks 10% of parents' combined income, whatever that is. It also offers other benefits for breastfeeding mothers (see ‘Express delivery’, left).

The NPA has found that child care remains a ‘mummy matter’, with women postdocs paying almost double the care costs that men pay. This has led the association's president, Alyson Reed, to speculate that more men must have access to partial child care from a spouse or relative.

Access to affordable child care opens the door to research-minded women. As Bertolotti puts it: “Knowing her child is well cared for frees her mind to focus on the science at hand.”

Virginia Gewin is a freelance science writer based in Portland, Oregon.

EXPRESS DELIVERY

As women choose to breastfeed for longer — citing numerous health benefits — the availability of dedicated areas for them to pump milk is increasingly an issue in the United States. Women often have to hunt for private spots to pump, such as vacant offices, service corridors or, most often, the toilet. Scandinavian culture is less restrictive. “There are few social barriers to sitting and nursing a baby during a seminar or at a lunch or even parking your buggy in the hallway,” says Jessica Marks, a biologist at the University of Bergen in Norway. And in Germany, women are allowed paid breastfeeding breaks — even if they have to travel 10 kilometres to reach their child.

V.G.

MOVERS

Albert-László Barabási, visiting scientist, Dana-Farber Cancer Institute, Harvard University, Boston, Massachusetts



2000–present: Emil T. Hofman Professor of Physics, University of Notre Dame, Notre Dame, Indiana

1999–2000: Associate professor, University of Notre Dame, Notre Dame, Indiana

1995–99: Assistant professor, University of Notre Dame

1994–95: Postdoc, IBM T. J. Watson Research Center, Yorktown Heights, New York

Walking through New York City during his years as a postdoc, Albert-László Barabási was intrigued not by the glitz and glamour, but by the notion of the invisible networks of cables and pipes necessary for modern life. Given that he trained as an engineer in Bucharest, Romania, such thoughts were not unusual for the Hungarian native.

But it is the connection between networks and another passion of his — chaos and fractals — that really underpins his scientific career. His fascination for fractals led him to Budapest to work with international fractal expert, Tamás Vicsek. Later, in New York, he married the ideas of the self-replicating patterns in fractals with the structure of networks.

Left to his own devices at IBM, Barabási began asking himself: “What the heck is a computer?” Realizing that everything from computers to electricity distribution to water pipelines is networked, he wondered why nobody in science had paid much attention to networks. “Networks must not be random, but we didn’t know anything about them,” he says. So he decided to fill in the gaps.

Timing was on his side. The appearance online of digital maps of the Internet and biological networks in the late 1990s became the foundations of his work. The result was two papers introducing the concept of scale-free networks. No matter what system he looked at, Barabási discovered that all networks are dominated by a few highly connected nodes or hubs.

Barabási decided to devote his full attention to networks, even though he lacked both funding and tenure. He credits this bold move with his ultimate success. “I thought this is going to be more important than anything I’ve done before, I can’t do it half-hearted or half-brained,” he says.

Now Barabási has his eyes on the next challenge: a theory of complexity. “I believe that if there will be a theory of complexity it will emerge in the next ten years,” he says. “How do I position myself to contribute to that goal?”

He is optimistic that his one-year move to the Dana-Farber Cancer Institute at Harvard University, will help him. As the field has advanced, Barabási says, it has become increasingly important to be close to experimental groups generating data.

For those who want to follow in his footsteps, Barabási offers one piece of advice: aim very high. “You’ll never reach that very high aim, but if you reach 75%, it’s still very good,” he says. “If you aim low, 75% gets you nowhere.” ■
Virginia Gewin

RECRUITERS & ACADEMIA

A level playing field

Funding pressures from governments have long helped to determine the direction of scientific research. But since the 1980s, the mantra ‘publish or perish’ has become increasingly pervasive. With more and more scientists competing for the limited space in the high-impact journals that institutions use to assess performance and award jobs or tenure, career progression seems an almost impossible challenge — especially for young scientists.

So, faced with such fierce competition, how does a postdoc gain a fellowship, or even a position in an institute? Grace Wong, chief scientific officer at ActoKine Therapeutics, a biotechnology company in Boston, says that apart from publishing in high-impact journals, young scientists can advance their careers by working with mentors who have strong international reputations (*Nature Biotechnol.* **23**, 265; 2005).

But working in a high-profile lab can bring its own problems. The level of competition can mean that younger scientists may have their work stamped on by senior colleagues. In the top laboratories, discoveries can lead to ugly situations, such as arguments over authorship or credit. And some principal investigators will assign two or three people to the same project to increase chances of success; the first to get the answer gets credit, a paper and

perhaps a job. Reputation then becomes a very delicate issue and objectivity in determining scientific output becomes distracting.

Ideally, other measures of success could be used. One suggestion has senior scientists being evaluated not solely on the number of papers they have published, but on how many people’s careers they have helped (see *Nature* **434**, 801; 2005). If this approach were embraced, the need to establish senior people in institutes to advise postdocs and students on scientific issues would become greater.

Other possible changes include creating equality within the laboratory instead of a hierarchy, so that each postdoc and PhD student has equal access to the mentor’s experience. And establishing clear demarcation for members of the same lab and between labs on a project could remove unnecessary competition, and allow everyone the chance to publish something.

Judging young scientists not just on publications but on other contributions would create a more welcome climate. But this won’t happen unless scientific managers are rewarded for creating such an environment. ■

Michael Edel is a postdoc at the Center for Genomic Regulation in Barcelona, Spain.

GRADUATE JOURNAL

The right path

Today, I am up to my neck in neuroscience, as next week I sit my final examination — after 12 months, two research projects, several essays and dozens of lectures. It has been a good year.

Last year I chose not to apply for a programme or scholarship in Germany, my home country. As a result, I ended up here in Oxford. But I think I prefer the approach taken by Anglo-American programmes when they assess candidates for a course. They tend to award positions on the basis of past research experience, a personal interview and the passion a candidate conveys in an interview. Germany, in contrast, tends to be highly bureaucratic. A stack of paperwork and written entry examinations come first. The personal interview is only the last step in the application procedure.

I was afraid of losing out before I reached the interview. Although I had achieved good results throughout my course, the marks were not exceptional. This was partly because medicine covers more than 40 subjects. How could one possibly be interested in all of them? I spent a lot of time in the lab concentrating on what really intrigued me. This fervour helped me to secure my current position. And, more importantly, it showed me that my style of studying is not wrong. ■

Tobias Langenhan is a first-year graduate student in neuroscience at the University of Oxford, UK.

Men sell not such in any town

In the realm of the senses.

Nalo Hopkinson

"Did you hear? Rivener has created a new fruit!"

"How dull. Her last piece was a fruit, too."

"Not like this one!" Salope said. She sat me at the table, murmuring the evening benediction as she did so. She draped my long sleeves artfully against the arms of the chair. She took my hat and veil, and hung them on the peg. She plucked the malachite pins from my hair, one by one. She shook the dark springing mass free, and refashioned it into a plait down my back. I endured as long as I could, then leaned back and stared up into her cool granite eyes.

"Tell me of Rivener's creation," I commanded her.

She came around to my side. She slipped her fingertips into the pockets of her white apron and composed herself for the tale. She stood quite straight, as was proper. My blood quickened.

"Rivener's previous fruit," she said, "only sang like a rainforest full of parrots; only enhanced the prescient abilities of those who ate it. This one is the pinnacle."

She stopped, although she didn't need breath. I felt a single drop of sweat start its slow trickle between my breasts. The heavy silks were stifling. "Stop dawdling. Tell me!"

She caught her bottom lip between gleaming teeth. She came and draped my sleeves into a second ritual form: the shape of mourning doves. I gritted my teeth. She continued: "It is the colour of early autumn, they say, and the scent lifting off its skin is a fine bouquet of virgin desire and dandy's sweat, with a top note of baby's breath. It fits in the palm, any palm. Its flesh is firm as a loving father's shoulder."

She stopped to dab at my face with a cutwork linen handkerchief from her pocket, and I nearly screamed. She resumed: "The fruit shucks off its own peel at a touch, revealing itself once only; to its devourer. A northern dictator burst into tears at the first taste of its pulp on his lips, and begged the forgiveness of his people."

"Poet and thrice-cursed child of a damned poet!" Her father too had played this game of stirring exalted cravings in me. I lifted the bodice away from my skin, fanned it to let air in. It wasn't enough.

Salope squatted in her sturdy black shoes, square at heel and toe. This exposed her strong thighs, brought her face level with my bosom. "I'm making you hungry, aren't I? Thirsty?"

"Bring me some water. No, wine."

"At once." She left the room, and returned with a sleek glass pitcher and a glass on a silver tray. The golden liquid was cold, and beaded the pitcher. Salope poured for me, tilted the glass to my lips. I tasted the wine. It was dry and dusty in my mouth. I turned my head away. "What does Rivener call this wonder?" I asked.

"The God Under the Tongue." Salope put the glass down on the table and took the appropriate step backwards. "There are one hundred and seventeen, limited edition, each one infused with her signature histamine."

"The one that makes the fingertips tingle?"

"The very same."

This heat! It distracted one so. "I wish to purchase one of these marvellous fruits."

"To taste it?"

"Of course to taste it! Bring me my meal."

"Instantly." She went. Returned with a gold dish, covered with a lid of sleekest bone. Fashioned from the pelvis of a whale; I knew this. She put the dish down, uncovered it. A fine

steam rose from it. "Here is your supper, Enlightened."

I picked up the golden spoon. "Contact the auction house."

Salope barely smiled. "I already have. It's too late. All one hundred and seventeen of 'The God Under the Tongue' are spoken for."

I slammed the spoon back down onto the table. "Tell them I will pay! Command Rivener to make another! Just one more!"

Salope looked down at the ground. When she returned her gaze to mine, she was serene. "It's too late, Enlightened. The Academy has decided. Rivener has been transmigrated to Level Sublime. She is beyond your reach."

"Machine."

"There is no need for insult, Enlightened."

"Go away."

Salope bowed, returned the spoon to my hand, and dissipated into black smoke. I preferred a pale rose mist, but Salope kept stubbornly reverting to black. It had been her father's favourite colour.

Perverse poet's child; how she could arouse the senses! Her father finally pushed me too far. I'd ordered him to dissolve himself permanently from my aura. I had grieved for two voluptuous years, then sought everywhere for his like. Nothing. Eventually, in desperation, I had summoned his daughter.

I am Amaxon Corazón Junia Principia Delgado the Third, and I bent over my meal and wept luxurious tears into my green banana porridge. It was a perfect decoction, and it now would not satisfy me. Only the poet's daughter, and her father before her, ever saw me so transported.

The room spoke. "Thank you, Enlightened. I consider myself well paid for today's session. Please recommend me to your acquaintances."

I would.

Nalo Hopkinson lives and works in Toronto, Canada. Her writing has received Honourable Mention in Cuba's Casa de las Americas Prize for fiction. She has just completed a new novel.



JACEY

Abstractions



FIRST AUTHOR

On page 412, Leslie Osborne and her colleagues report the results of experiments to investigate whether variability in fine

movements is the result of sensory or motor errors. The group trained monkeys to visually track spots of light moving on a monitor, and measured their eye movements. Then they analysed the variability in the movement and found that it could be explained by errors in the sensory estimate of the target, as opposed to noise in the motor system itself.

Osborne, a neurobiology postdoc at the University of California, San Francisco, has a background in biophysics and a postdoc in neuroscience. This range of skills helped her to serve as a conduit between Steve Lisberger, her biologist principal investigator, and Bill Bialek, her theoretical physics mentor. *Nature* found out more about her bridging role.

How did you cross from physics to biology?

At Berkeley, where I did my PhD, I spent a lot of my time in a laser-physics lab building a device to measure small movements. I used this to study how tiny hairs on crickets help the insects to detect predators.

What were the roles of the other authors?

The work grew out of conversations I had with Bill about variability in tracking eye movements and whether we could look for evidence that any variability seen could be explained by noise in the sensory system. I did the experiments. Bill provided the theoretical framework and I analysed all the data. Steve is great at making you define your idea, until you think that your conclusions are not just a mathematical construct. He was our resident sceptic.

What was it like to work cross-discipline?

The biggest challenge was that Bill and Steve speak very different languages. Writing the paper was tricky — that's when you see how different the languages of the two disciplines can be. Bill is comfortable expressing himself in mathematics. Biologists and neuroscientists don't tend to communicate this way. We took most of the maths out of our paper — there were five pages of maths that didn't make it into the final submission.

Why does biology need more theory?

Because it's theory that drives new experiments. If you have an overview on how a system might work, it allows you to ask more shrewd questions that might prove or disprove your hypothesis.

What's next?

To see if our finding generalizes to other motor systems. Eventually, I would like to track neural activity along with behaviour to watch the brain's translation of vision to action emerge. ■

MAKING THE PAPER

Shehzaad Kaka
and Fred Mancoff

Power from some devices can be more than the sum of their parts.

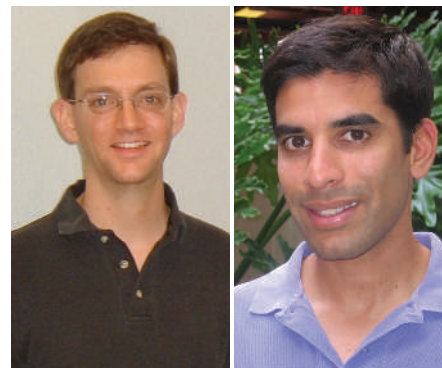
Fireflies flashing in unison...pedestrians marching over bridges and falling in step with each other... Nature has a way of getting things in synch. So it should come as no surprise that two separate groups — one academic, one industrial — simultaneously pursued such effects in tiny devices (see pages 389 and 393).

The starting point for both groups was an earlier finding that nanosized magnetic oscillators can generate microwaves. By combining multiple devices, the teams hoped to generate sizeable amounts of microwave power. To test out this idea, they set about linking two devices. This, they hope, will serve as the first step towards building a simple wireless circuit, which could pave the way to faster computation.

Selecting and assembling the building blocks for their devices posed difficult problems for both groups. The fabrication process was "fairly complicated", says Shehzaad Kaka, first author on the paper from the public effort, based at the US National Institute of Standards and Technology (NIST) laboratory in Boulder, Colorado. Fred Mancoff, first author with the commercial group at Freescale Semiconductor in Chandler, Arizona, agrees.

The devices are made up of several layers of thin-film materials, and patterning the tiny magnetic oscillators meant using electron beam lithography to resolve the features between neighbouring oscillators. They had to find ways to monitor how the devices worked, both as individuals and when the two were hooked up together. "At first, we weren't really set up for measuring these high-frequency signals," Mancoff says.

As the two groups moved from studying individual oscillators to two coupled devices, they took slightly different approaches. For each set of measurements, the NIST group



Fred Mancoff (left) and Shehzaad Kaka.

kept its two oscillators in the same position. The Freescale researchers focused instead on what happened as they changed the distance between the two devices.

Despite the different approaches, both groups quickly saw the same thing: at some point, the electrical output signal from each of the two devices began oscillating at the same frequency. "It was something we were hoping to get," says Kaka.

Both teams were surprised at how quickly they got a result. "From the first step, there was a fairly strong effect going on as a result of the spacing," Mancoff says. "I knew that there was probably some interesting physics going on."

Each of the two groups was aware that the other was doing similar research — not surprising as the two first authors had shared an office at NIST when Kaka was a PhD student and Mancoff was a postdoc. So, having validated their work with more experiments, they both scrambled to submit a paper. "Once we found that Freescale was also intending to publish, that definitely motivated us to finish up, write it up and send it in for publication," says Kaka.

Although neither group demonstrated that the devices in tandem can generate sizeable amounts of power, both groups aim to combine more devices. Thanks to the potential applications of the devices, the groups expect to see even more competition in this field. And they might even see more developments that seem to happen in synch. ■

QUANTIFIED CHILE

A numerical perspective on *Nature* authors.

In Chile, pursuing science at a regional university such as the Pontifical Catholic University of Valparaíso is not easy, says Marco Cisternas. The size and influence of competing universities in Santiago, the country's capital, means that funding from government agencies is hard to come by for scientists in smaller cities.

Cisternas is lucky to have had full support for his research, but really, he says, the only way to work is to team up with other national universities and with foreign colleagues. Cisternas is a member of a multidisciplinary team with scientists from the United States, Japan, India and Indonesia, as well as Chile. The group works on earthquakes, and its most recent research maps out giant-earthquake recurrence over the past two millennia in southern-central Chile (see page 404).

10 authors of papers published in *Nature* so far this year are working in Chile (total number of published authors = 3,984).

3 *Nature* papers published this year have contributing authors working in Chile (total number of published papers = 599).

14 Submissions to *Nature* this year have come from Chile.

3 authors from the Pontifical Catholic University of Valparaíso have research published in *Nature* this week.